

Magisterarbeit

Internationales Informationsmanagement

Entwicklung und Evaluierung von
Varianten des multimodalen Dialogs für
Fahrerinformationssysteme im Kraftfahrzeug

Isabelle Peters

Universität Hildesheim

Erstgutachter: Dr. Thomas Mandl

Zweitgutachterin: Prof. Dr. Christa Womser-Hacker

Institut für Angewandte Sprachwissenschaft

Universität Hildesheim

Betreuer der Robert Bosch GmbH:

Dr. Andreas Korthauer

Hildesheim, 15. Januar 2005

Zusammenfassung

Diese Arbeit behandelt die Weiterentwicklung eines multimodalen Dialogsystems für die Navigationszieleingabe in einem Kraftfahrzeug. Sowohl die graphische Oberfläche als auch die Sprachbedienung wurden - insbesondere im Hinblick auf eine effektive und effiziente Listenauswahl - verbessert und im Anschluss daran in einem umfassenden Benutzertest evaluiert.

Abstract

This paper illustrates the further development of a multimodal dialogue system for entering destinations in a satellite navigation system within motor vehicles. Both the graphical user interface and the speech dialogue have been improved (with particular reference to creating an effective and efficient selection list) and evaluated in an extensive user test.

Inhaltsverzeichnis

Zusammenfassung	II
Vorwort	VI
1 Einleitung	1
1.1 Motivation	1
1.2 Aufgabenstellung	3
1.3 Aufbau dieser Arbeit	4
2 Grundlagen	5
2.1 Fahrerinformationssysteme	5
2.2 Sprachbedienbare Dialogsysteme	8
2.3 Multimodale Dialogsysteme	13
2.3.1 Multimodalität	13
2.3.2 Vorteile eines multimodalen Dialogsystems	13
2.3.3 Anforderungen an multimodale Dialogsysteme	14
2.3.4 Multimodale Dialogsysteme im Kraftfahrzeug	16
3 Beschreibung des Dialogsystems	18
3.1 Die einzelnen Komponenten und ihr Zusammenspiel	18
3.2 Der Dialog mit dem Testsystem	24
3.2.1 Bedienung über die graphische Oberfläche	24
3.2.2 Sprachbedienung	32
3.2.3 Multimodale Bedienung	39
4 Entwicklung der Variante	42
4.1 Durchführung der explorativen Studie	42
4.1.1 Struktur und Aufgabenstellung	42

4.1.2	Ergebnisse	44
4.2	Weiterentwicklung des Dialogs	46
4.2.1	GUI-Variante	46
4.2.2	Sprachdialog-Variante	50
5	Benutzertest	59
5.1	Fragestellung und Hypothesen	59
5.2	Durchführung des Benutzertests	62
5.3	Ergebnisse	66
5.3.1	Überprüfung der Hypothesen	66
5.3.2	Weitere Ergebnisse	74
6	Fazit und Ausblick	77
 Anhang		
A	Buchstabieren	81
A.1	Buchstabieralphabeten	81
A.2	Verwendete Wörter	82
B	Dialogabläufe des Testsystems	83
B.1	Legende	83
B.2	Hauptmenü	84
B.3	Listenansicht	85
B.4	Filteransicht	86
C	Explorative Studie	87
D	Benutzertest	91
D.1	Die Versuchsanweisungen	91
D.2	Die Aufgabenstellung	93
D.3	Der Fragebogen	96
D.4	Der Vorabfragebogen	102
Tabellenverzeichnis		105

Abbildungsverzeichnis	106
Literaturverzeichnis	108

Vorwort

Die vorliegende Magisterarbeit entstand in Zusammenarbeit mit der Forschungsabteilung FV/SLI (Informations- und Systemtechnologie) der Firma Robert Bosch GmbH in Stuttgart. Daher möchte ich mich zunächst bei den Herren Lars Biester und Andreas Kynast für die Herstellung des Kontaktes und die Ermöglichung dieser Arbeit bedanken.

Mein besonderer Dank gilt meinem Betreuer Herrn Dr. Andreas Korthauer, der konstruktive Kritik und Ratschläge dort anbrachte, wo sie notwendig waren, ohne mir jedoch meine Selbstständigkeit zu nehmen. In diesem Zusammenhang danke ich auch Herrn Dr. Ulrich Schweiger und Frau Miriam Hausmann für spontane Hilfestellungen und anregende Diskussionen. Allen Mitarbeitern von FV/SLI3 möchte ich dafür danken, dass Sie durch ihre Arbeitsmoral und Kollegialität zu einer Atmosphäre beigetragen haben, in der ich mich gern aufhielt und effektiv arbeiten konnte.

Nicht unerwähnt lassen möchte ich die Teilnehmer des Psychologen-Forums, die einen interessanten fachlichen Austausch über Evaluierungskonzepte und Auswertungsmethoden ermöglichten.

Nicht zuletzt danke ich all jenen, die die Mühen des Korrekturlesens auf sich nahmen. Da diese Magisterarbeit das Ende meines Studiums markiert, möchte ich mich an dieser Stelle auch bei meiner Familie und meinen Freunden für ihre langjährige Unterstützung bedanken.

Stuttgart, im Januar 2005

Isabelle Peters

Anmerkung der Verfasserin:

In dieser Arbeit wird wegen der besseren Lesbarkeit nur die männliche Form verwendet, wenn in einem allgemeinen Sinne von Personen die Rede ist.

1 Einleitung

1.1 Motivation

Radio, CD-Wechsler, Navigationssystem, Einparkhilfe, Telefon, Bordcomputer, Fernseher - die Liste ließe sich beliebig fortführen. Immer mehr Autofahrer legen großen Wert auf derartige Ausstattungen zur Information, Kommunikation und Unterhaltung im Auto. Und die Automobilindustrie erfüllt diese Kundenwünsche in zunehmendem Maße. Schon heute wird kaum ein Neuwagen ohne Freisprechanlage, Radio und CD-Wechsler verkauft. Bei höherklassigen Fahrzeugen gehören Navigationssystem und Bordcomputer häufig bereits zur Grundausstattung.

Doch mit der steigenden Zahl der im Auto verfügbaren Funktionen wird ihre Bedienung nicht einfacher - im Gegenteil. Große Bildschirme, Multifunktionslenkrad und andere Kontrolleinheiten sollen dem Fahrer zwar eine einfache und bequeme Bedienung ermöglichen, sind jedoch häufig auch der Grund für Unsicherheit. WOMIT bedient man WAS und vor allen Dingen WIE? Viele, insbesondere ältere Fahrer fühlen sich durch die Vielzahl der Funktionen und Möglichkeiten überfordert. Darüber hinaus stellt die Bedienung von Informations- und Kommunikationssystemen während der Fahrt ein nicht zu unterschätzendes Sicherheitsrisiko dar. Wenn ein Großteil der Konzentration auf den Umgang mit einem solchen System und nicht mehr auf das Verkehrsgeschehen gerichtet ist, sind schnelle Reaktionen in Gefahrensituationen nicht mehr möglich. Zwei nicht unbedeutende Punkte sind dabei die Blickabwendung (beispielsweise zum Überprüfen des Displays) und das Loslassen des Lenkrades zumindest mit einer Hand (z.B. zur Bedienung des Radios). Nicht zu Unrecht wurde in Deutschland zum 1. April 2004 das Telefonieren ohne Freisprechanlage während der Fahrt unter Strafe gestellt¹. Daher ist

¹ Handyverbot am Steuer, vgl. §23, Absatz 1b der Straßenverkehrsordnung für die Bundesrepublik Deutschland: „Dem Fahrzeugführer ist die Benutzung eines Mobil- oder Autotelefons untersagt, wenn er hierfür das Mobiltelefon oder den Hörer des Autotelefons aufnimmt oder hält.[...]“[BMV04]

es Ziel der Automobilindustrie, den Autofahrern ein System zur Verfügung zu stellen, dessen Bedienung klar und einfach, natürlich und bequem, aber vor allen Dingen auch sicher ist.

Die Bedienart spielt dabei eine wesentliche Rolle. Ist nur eine Bedienung über die graphische Oberfläche (im Folgenden kurz GUI, Graphical User Interface, genannt) möglich, steigt das Gefahrenpotenzial, da die Augen häufig von der Fahrbahn abgewendet werden. Außerdem muss meist auch die Hand vom Lenkrad genommen werden, um die Bedienung vorzunehmen (z.B. um den Touchscreen zu berühren oder ein Kontrollrad zu drehen). Ein Multifunktionslenkrad stellt bereits eine bedeutende Verbesserung dar, da hier über zahlreiche Tasten und Schalter am Lenkrad viele Funktionen bedient werden können. Aufgrund der begrenzten Anzahl der Tasten auf dem Lenkrad und der immer weiter steigenden Komplexität der Informationssysteme können jedoch bereits heute nicht mehr alle Funktionen über das Lenkrad bedient werden. Darüber hinaus sind auch hier Blickabwendungen nötig.

Sprachbedienung scheint eine Lösung für dieses Problem zu sein. Dabei können die Hände am Lenkrad und der Blick auf der Straße bleiben. Allerdings gibt es auch hier Situationen, in denen eine visuelle Kontrolle notwendig ist. Beim Navigationssystem ist dies beispielsweise bei der Wahl des Zielortes der Fall, wenn Homophone² vorhanden sind. Möchte der Fahrer nach Ahlen, könnte es passieren, dass das System ihn nach Aalen führt, da die Aussprache der beiden Orte gleich ist. Um dies zu verhindern, sollte der Fahrer wenigstens die Möglichkeit bekommen, seine akustische Eingabe visuell zu kontrollieren. Des Weiteren kann es passieren, dass der Benutzer bei einem sprachbedienbaren System an eine Stelle gelangt, an der er nicht weiter kommt, sei es aufgrund eines Sprachdialektes oder einer schlechten Erkennung aufgrund von Umgebungsgeräuschen. Darüber hinaus kann es vorkommen, dass selbst die definierten Rückfallebenen³ eines Sprachdialogs (z.B. das Buchstabieren) nicht zum Ziel führen. In solch einem Fall muss der Benutzer seine Eingabe auch über eine andere Modalität als die Sprache vornehmen können.

² Homophon, [gr.] Wort, das mit einem anderen gleich lautet, aber verschieden geschrieben wird (z.B. Lehre - Leere) [Dud01]

³ Rückfallebene, innerhalb eines Sprachdialogs festgelegter Dialogzustand, der möglichst einfach gestaltet sein sollte, um dem Benutzer die Möglichkeit zu geben, auch im Falle von auftretenden Problemen eine Eingabe machen zu können.

Die vorangehenden Gedanken sollten klargemacht haben, dass im Auto einzig eine multimodale Bedienung sinnvoll erscheint. Dabei sollte die Wahl der Bedienart allerdings immer beim Benutzer des Systems liegen. Der Fahrer oder sein Beifahrer kann - je nach Situation und Vorliebe - unimodal, multimodal oder sogar mehrere Applikationen parallel über verschiedene Modalitäten bedienen. Auf diese Art und Weise sind sowohl Benutzerfreundlichkeit als auch Verkehrssicherheit gewährleistet.

1.2 Aufgabenstellung

Ein wichtiger Forschungsbereich der Robert Bosch GmbH ist das gesamte Umfeld der Mensch-Maschine-Schnittstelle für Fahrerassistenz- und -informationssysteme. Dazu gehören unter anderem auch Navigationssysteme. In diesem Forschungsbereich wurde im Rahmen eines Robert-Bosch-internen Projekts ein Prototyp für die Gestaltung des Zieleingabe-Dialogs eines Navigationssystems entwickelt. Dieser Prototyp ist multimodal (mit Sprache und/oder über die graphische Oberfläche) bedienbar und soll der Evaluierung multimodaler Bedienkonzepte dienen. Innerhalb des Zieleingabe-Dialogs ist dabei die Auswahl des Zielortes aus einer langen Liste ebenso möglich wie die Einschränkung der Liste mittels einer Filterbedingung (Buchstabieren).

Ziel dieser Arbeit ist es nun zunächst, die bereits realisierte Testversion zu bewerten. Dabei sollen Schwachstellen aufgedeckt und mögliche Verbesserungsvorschläge erdacht werden. Dies beinhaltet Fragestellungen wie: Welche Funktionen werden von den Versuchspersonen nicht verstanden/benutzt? Gibt es Situationen, die vom Sprachdialog nicht abgefangen werden? Wie funktioniert das Buchstabieren mit Eigennamen?

Darauf aufbauend sollen Varianten zu dem bestehenden Dialog entwickelt werden. Diese Varianten sollen ebenso wie die Ausgangsversion multimodal bedienbar sein und Listenauswahl sowie Filterbedingung enthalten. Zur Entwicklung der Varianten sind sowohl Änderungen in der Java-basierten Ausgangsversion als auch im XML-Skript für den Sprachdialog und in der dazugehörigen Grammatik für den Spracherkenner notwendig. Die ausgearbeiteten Varianten sollen schließlich in einem Benutzertest evaluiert werden. Dabei sollen neben Effektivität und Effizienz der entwickelten Varianten auch die Verhaltensweisen der Benutzer im Hinblick auf Modalitätsverwendung bzw. -wechsel begutachtet werden. Ein wichtiger Aspekt bei der gesamten Entwicklung ist die Berücksichtigung des Kraftfahrzeug-Umfelds.

1.3 Aufbau dieser Arbeit

In Kapitel 2 werden die Grundlagen von Interaktionssystemen zwischen Mensch und Computer dargestellt. Dabei liegt das Hauptaugenmerk auf Systemen im Kraftfahrzeug, da die zu entwickelnden Varianten dort ihre Anwendung finden werden. Zunächst wird daher auf Fahrerinformationssysteme im Allgemeinen, anschließend gezielt auf sprach- und multimodal bedienbare Systeme eingegangen.

Kapitel 3 enthält eine Beschreibung des realisierten Testsystems. Sowohl der Java-Teil als auch das XML-basierte Dialogskript werden erläutert und das Zusammenspiel zwischen Spracherkenner, Testsystem und Sprachsynthese erklärt.

Darauf aufbauend wird in Kapitel 4 die Weiterentwicklung des Dialogs dargestellt. Nach einer kurzen Skizzierung der durchgeführten explorativen Studie werden die Änderungen in der Grammatik für den Spracherkenner, im Java-Code und im XML-Skript genannt. Zudem werden die Beweggründe für diese Änderungen verdeutlicht.

In Kapitel 5 werden der Aufbau des Benutzertests, seine Durchführung sowie die Ergebnisse beschrieben. Einzelne Punkte werden kritisch vor dem Hintergrund der Übertragbarkeit auf alle Autofahrer diskutiert.

Das abschließende Kapitel 6 beinhaltet eine Zusammenfassung der Ergebnisse und einen Ausblick auf weiterführende Studien und Forschungsmöglichkeiten.

2 Grundlagen

2.1 Fahrerinformationssysteme

Sowohl innerhalb der Automobilindustrie als auch in der Literatur herrscht keine Einigkeit über die Definition eines Fahrerinformationssystems (im Folgenden FIS genannt). Beispielsweise dokumentiert die AUDI AG ihr FIS, das in Abbildung 2.1 zu sehen ist, folgendermaßen:

„Das Fahrerinformationssystem im Kombi-Instrument ist Anzeige- und Informationszentrum des Fahrzeugs. Es informiert über den aktuellen Betriebszustand des Wagens. Hierzu gehören beispielsweise die Meldungen der Auto-Check-Control und - je nach Ausstattung - Radio-, Telefon- und Navigationsdaten.“

(<http://www.audi.com/jsp/crossfeatures/lexicon/lexiconDetails.jsp?lexiconId=14076&lexiconChar=F&domainId=2&languageId=3>)



Abbildung 2.1: Fahrerinformationssystem der AUDI AG (Quelle s.o.)

Bei der Robert Bosch GmbH wird die Informationsdarstellung im Zusammenhang mit FIS zumeist etwas anderes verstanden: Die Informationen werden nicht nur im Kombiinstrument, sondern zusätzlich auch im Mitteldisplay angezeigt. Ein Beispiel hierfür ist

das Navigationssystem „TravelPilot“ von Bosch-Blaupunkt, bei dem neben der Fahrtrichtungsanzeige im Kombiinstrument zusätzlich Karteninformationen o.ä. im Mitteldisplay dargestellt werden können.

Geschieht die Informationsdarstellung hauptsächlich zur Gefahrenwarnung und Unfallprävention, wie es z.B. beim ACC¹ der Fall ist, ist auch die Bezeichnung Fahrerassistenzsystem häufig anzutreffen. Einige Fahrerassistenzsysteme können sogar aktiv in das Verkehrsgeschehen eingreifen. Beispielsweise kann das ACC bremsen, um einen Auffahrunfall zu verhindern. Der Trend geht dahin, Fahrerassistenz- und Fahrerinformationssysteme miteinander zu vernetzen.

Lyle Saxton, der Geschäftsführer des „Office of Safety and Traffic Operations Research and Development“ der US-amerikanischen FHWA², formulierte 1988 das Ziel, ein „Advanced Driver Information System (ADIS)³“ zu entwickeln. Darunter verstand er Folgendes:

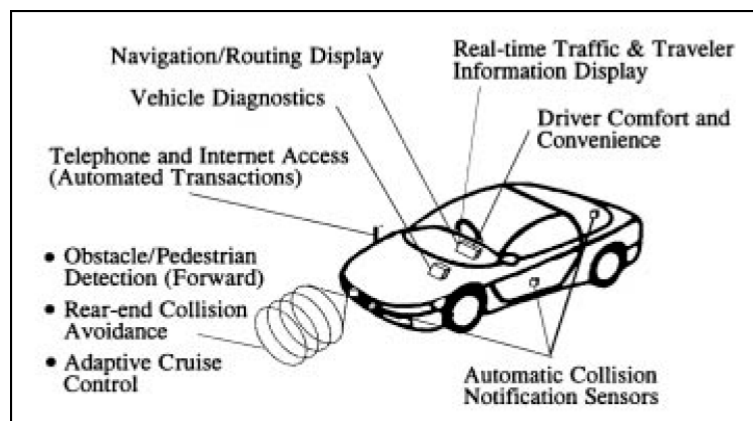


Abbildung 2.2: Advanced Traveler Information Systems Configuration. [FHW98, S. 1]

„A comprehensive in-vehicle system which provides a broad range of real-time information to improve motorist safety, convenience and traffic operations.“ (Lyle Saxton (FHWA) 1988, zitiert nach [Top90, S. 364])

¹ ACC: *Adaptive Cruise Control*. System zur automatischen Abstandsmessung und -einhaltung.

² FHWA: *Federal Highway Administration*. Behörde des U.S. Department of Transportation, zuständig für Fragen rund um Kraftverkehr und -fahrzeuge, vergleichbar mit dem Verkehrszweig des Bundesministeriums für Verkehr-, Bau- und Wohnungswesen.

³ Das Projekt „ADIS“ existiert auch heute noch, wurde allerdings umbenannt in „ATIS“ = Advanced Traveler Information System.

Abbildung 2.2 zeigt ein Beispiel für ein derartiges Fahrerinformationssystem.

Diese Definition aus dem Jahr 1988 lässt sich auch auf heutige FIS gut anwenden. Die frühere Trennung zwischen Systemen, die den Fahrzeugzustand (Geschwindigkeit, Drehzahl, Kilometerstand, Warnlampen) anzeigen und solchen, die Informationen zur Verkehrslage (Navigation, Verkehrsfunk) bereitstellen oder der Unterhaltung (Radio, CD, MP3) dienen, löst sich immer mehr auf. Wie in Abbildung 2.2 am Beispiel des ATIS gut zu erkennen ist, beinhaltet ein FIS alle möglichen Funktionen wie z.B. Fußgängererkennung, Telefon oder Navigationssystem.

Zur besseren Übersicht lassen sich diese Funktionen drei Bereichen zuordnen (vgl. auch Abbildung 2.3):

1. Technologien zur Gefahrenwarnung und Fahrtunterstützung: ACC, Hindernis- und Fußgängererkennung, Umfelderkennung, LDW⁴, Schilderkennung, Einparkhilfe, Tempomat
2. Vorrichtungen zur Fahrerinformation: Navigation (Zielsuche und -führung), Verkehrsfunk und Reiseinformationen, Wetterbericht, „Gelbe Seiten“, Informationen über den Fahrzeugzustand (Geschwindigkeit, Drehzahl, Tankfüllung)
3. Kommunikations- und Unterhaltungs-Funktionen: Kommunikationselemente (Mobiltelefon bzw. Freisprechanlage, Fax, E-Mail), Vorrichtung zur Abwicklung von Bankgeschäften, Hotel- oder Restaurantreservierungen, Unterhaltungselektronik (Radio, CD, MP3, Fernsehen)

Die oben genannten Einzelsysteme gehören, einzeln oder auch zu einem FIS zusammengeschlossen, immer häufiger zur Ausstattung in neuen Fahrzeugen. Dabei gab oft die Automobilindustrie den Anstoß für die Entwicklung derartiger Informationssysteme, um den Bedürfnissen der Verbraucher, d.h. der Autofahrer, entgegenzukommen.

⁴ LDW: *Lane Departure Warning*. Warnsystem, das anhand der Leitplanken und Seitenstreifen die Fahrspur erkennt und den Fahrer vor dem unbeabsichtigten Verlassen der Spur warnt.

Basic Collision Warning Technologies: Adaptive Cruise Control; Rear-End Collision Avoidance; Obstacle/Pedestrian Detection.

Basic Traveler Information Devices: Navigation/Routing; Real-Time Traffic and Traveler Information; Automatic Collision Notification.

Driver Convenience Devices: Comfort and Convenience (Cellular Telephones, Fax, E-mail); Vehicle Diagnostics; Automated Transactions.

Abbildung 2.3: Drei Funktionsbereiche des Advanced Traveler Information System [FHW98, S. 1]

2.2 Sprachbedienbare Dialogsysteme

Joseph Weizenbaums „Eliza“-System aus dem Jahr 1966, von Andreas Kellner als „Urvater aller Dialogsysteme“ [Kel01, S. 484] bezeichnet, war aus heutiger Sicht alles andere als ein sprachbedienbares Dialogsystem. Eliza ähnelte einem Psychotherapeuten, der auf Aussagen wie „Ich habe ein Problem mit meinem Vater.“ mit „Warum sagen Sie, haben Sie ein Problem mit Ihrem Vater?“ geantwortet hat. Dabei konnte die „Sprache“ allerdings nur in geschriebener Form ein- und ausgegeben werden und Elizas Sprachverstehen und Dialogverhalten beschränkte sich auf simple Schlüsselworterkennung⁵.

Heutige sprachbedienbare Dialogsysteme sind in ihrer Architektur allein deswegen viel komplexer, weil sie akustische Spracheingaben zulassen. Ihr Einsatz ist allerdings nach wie vor meist auf ein bestimmtes Anwendungsgebiet beschränkt. Als bekannte Beispiele sollen hier die Automatische Fahrplanauskunft der Deutschen Bahn AG oder auch das Bundesligainformationssystem „Berti“ der Sympalog Voice Solutions GmbH genannt werden. Letzteres gewann den Voice Award 2004⁶.

Die Architektur eines Sprachdialogsystems ist in Abbildung 2.4 beispielhaft dargestellt. Deutlich zu erkennen ist die Bedeutsamkeit jeder einzelnen Komponente des Dialogsystems. Nur durch das Zusammenspiel mehrerer intelligent gestalteter Einzelsysteme

⁵ Eliza System: Für detailliertere Erklärungen zu Elizas Systemverhalten vgl. <http://de.wikipedia.org/wiki/ELIZA>.

⁶ Voice Award: Die Initiative VOICE BUSINESS prämiiert mit dem Voice Award die besten deutschsprachigen Sprachapplikationen. Der erste Voice Award wurde 2004 vergeben (vgl. hierzu [Voi04]).

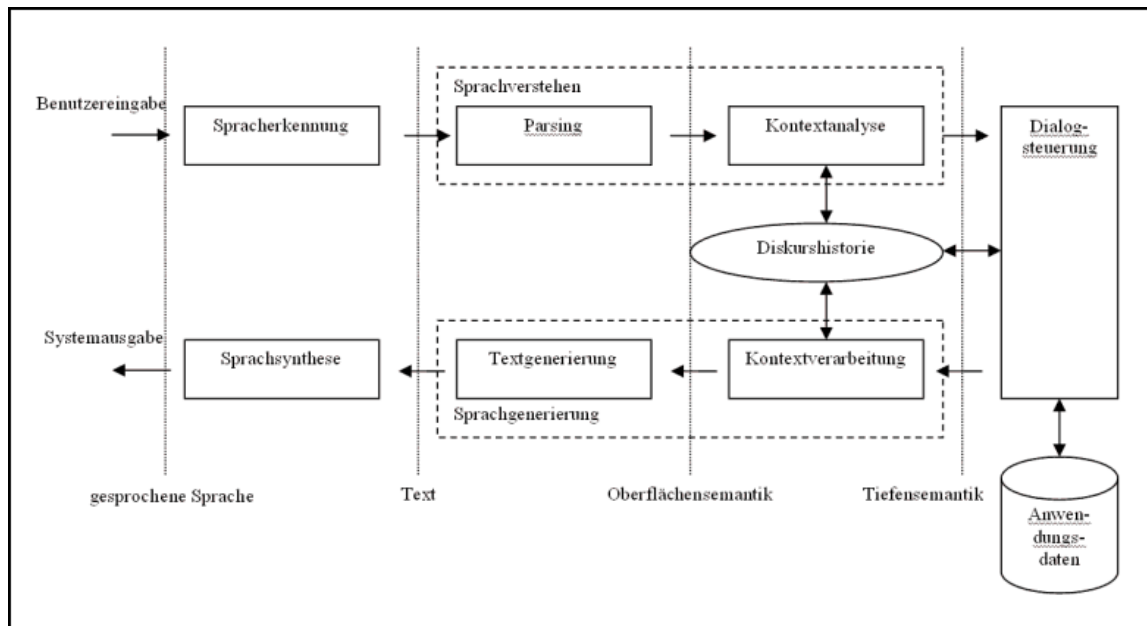


Abbildung 2.4: Architektur eines natürlichsprachlichen Dialogsystems [Kel01, S. 486]

kann ein Sprachdialogsystem erfolgreich mit Benutzern kommunizieren und deren Bedürfnisse befriedigen. Bereits das mangelhafte Design einer einzelnen Komponente führt zum Fehlschlagen des Dialogs bzw. zu mangelhafter Akzeptanz des Dialogssystems. Im Folgenden sollen die einzelnen Komponenten kurz skizziert werden.

Benutzereingabe

Die meisten Sprachdialogsysteme ermöglichen heutzutage eine akustische Benutzereingabe. Dabei soll zwischen vier Arten von Sprachdialogsystemen unterschieden werden. Diese vier Arten variieren im Grad der Restriktionen an die Benutzereingabe.

Die wohl älteste Form sind die *Kommando-basierten Systeme*, die den Benutzer auf eine kleine Anzahl an möglichen Wörtern einschränken (z.B. „Wenn Sie die Fahrplanauskunft möchten, sagen Sie 1, möchten Sie Wetterinformation, sagen Sie 2“ usw.). Sie sind jedoch eher die Seltenheit (vgl. hierzu [Pra04, S. 10]). Bereits weit weniger restriktiv sind *natürlichsprachliche Systeme*, die *kurze Phrasen als Benutzereingabe* zulassen. Die Benutzer können frei sprechen, sofern ihre Eingabe den innerhalb des Systems festgelegten Grammatikregeln entspricht; sogar Füllwörter sind möglich. Dennoch weisen diese Systeme weit weniger Flexibilität auf als *natürlichsprachliche Systeme mit offenem Vokabular*. Hier werden bereits Besonderheiten der natürlichen Sprache wie beispielsweise El-

lipsen⁷, Anaphern⁸ oder unbeendete Sätze bedacht. Allerdings wird davon ausgegangen, dass selbst diese Art des Dialogs vom Sprecher noch geplant bzw. durchdacht ist. Daher werden diese Systeme in der Literatur auch häufig als „Planned Speech Systems“ bezeichnet. Damit sollen sie von den *spontansprachlichen Systemen* abgegrenzt werden. Diese können, neben den oben bereits erwähnten „ungrammatischen“ Phänomenen auch Korrekturen und Fehlsätze und insbesondere auch nichtverbale Äußerungen wie Lachen, Husten und Schmatzen verarbeiten. Schukat-Talamazzini spricht hier bei den Benutzern von einer „minderen Sorgfalt der Artikulation“ [Sch95, S. 336f.]

Spracherkennung

Die Spracherkennung, auch häufig mit ASR (Automatic Speech Recognition) abgekürzt, transformiert das akustische Signal des Benutzers in Text. Sie basiert zumeist auf statistischen Modellen, wobei sich die so genannten Hidden Markov Modelle (HMM) als zentrales Element durchgesetzt haben. HMM modellieren die Wahrscheinlichkeit, mit der ein Abschnitt eines Sprachsignals ein bestimmtes Phonem⁹ darstellt. Damit lässt sich für ein gegebenes Sprachsignal die wahrscheinlichste Phonemfolge ermitteln, so dass mit Hilfe eines Lexikons mögliche Wörter abgeleitet werden können.¹⁰

Ein wichtiger Aspekt bei der Spracherkennung ist die Sprecherunabhängigkeit. Das System muss von jedem Benutzer ohne vorheriges Training unmittelbar verwendet werden können. Darüber hinaus hat das Vokabular einen entscheidenden Einfluss auf das Erkennungsergebnis. Je kleiner das Vokabular, desto besser werden im Allgemeinen die Erkennungsraten des Systems. Allerdings kann selbst ein kleines Vokabular die Erkennungsschwierigkeiten bei phonetisch verwechselbaren Äußerungen nicht mindern. Zudem muss bedacht werden, dass ein kleines Vokabular den Benutzer deutlich einschränkt und das System somit unflexibel wird. Weitere Problemstellungen bei der Spracherkennung wie beispielsweise den Lärm, der insbesondere auch in der Automobil-Umgebung auftritt, nennen Bernsen et al.:

⁷ Ellipse, [gr.-lat.] Ersparung, Auslassung von Redeteilen, die für das Verständnis entbehrlich sind, z.B. „der [Täter] oder die Täter sollen sich melden“ [Dud01].

⁸ Anapher, [gr.-lat.] zurückweisendes Element eines Textes, z.B. „Die Frau... Sie war sehr elegant.“ [Dud01].

⁹ Phonem, [gr.] (Sprachw.) kleinste bedeutungsunterscheidende, aber nicht selbst bedeutungstragende sprachliche Einheit [Dud01]

¹⁰ Für eine detaillierte Beschreibung der Funktionsweise von Hidden Markov Modellen vgl. [Sch95, Kapitel 5]

„Speech recognition must cater for extra-linguistic noise and other phenomena, such as that the speech rate varies over time, the speech signal is mixed with environmental noise from other people speaking, traffic and slamming doors, the pronunciation varies with the speaker, and speech from different participants may overlap, for instance with the system’s utterances.“ [Ber98, S. 53]

Sprachverstehen

Die Komponente für das Sprachverstehen wird zumeist als NLU (Natural Language Understanding) bezeichnet. Die NLU hat das Ziel, der erkannten Wortfolge eine Semantik zuzuordnen, d.h. die Bedeutung der Worte zu erkennen. Mittels Parsen, also der syntaktischen Analyse der Wortfolge, wird ein Parse-Baum erstellt, in dem die Funktion jedes einzelnen Wortes im Satz abgebildet ist. Aus diesem Parse-Baum kann daraufhin mit Hilfe der Kontextanalyse die Bedeutung des Satzes interpretiert werden. An dieser Stelle treten die oben erwähnten Probleme mit Ellipsen oder ähnlichen Sprachphänomenen auf. Hierzu schreibt Michael F. McTear:

„Spoken language introduces a further problem in that the output from the speech recognizer will often not have the form of a grammatically well-formed string that can be parsed by a conventional language understanding system. Rather it is likely to contain features of spontaneous speech, such as sentence fragments, after-thoughts, self-corrections, slips of the tongue, or ungrammatical combinations.“ [McT02, S. 107]

Um die genannten Probleme zu umgehen bzw. trotz ihres Auftretens ein robustes Sprachverstehen zu ermöglichen, wird häufig auf partielles Parsen zurückgegriffen. Falls also eine vollständige Interpretation der Eingabe nicht möglich ist, werden nur einzelne bedeutungstragende Abschnitte der erkannten Wortfolge untersucht. Darüber hinaus werden Informationen aus dem bisherigen Dialogverlauf und domänenspezifisches Wissen für die Interpretation herangezogen.

Dialogsteuerung

Die Dialogsteuerung, häufig auch als Dialogmanager bezeichnet, ist das Kernstück eines Dialogsystems. Hier wird, basierend auf den Ergebnissen der NLU, die Äußerung des Benutzers auf ihre Intention hin interpretiert und eine angemessene Systemreaktion ausgewählt. Beispielsweise könnte das System bei unzuverlässig verstandenen Eingaben oder mehreren möglichen Interpretationen nachfragen, um eine höhere Sicherheit bzw. Disambiguierung zu erreichen.

Dialogsysteme werden danach unterschieden, bei wem die Initiative während des Dialogverlaufs liegt. In der Literatur herrscht Einigkeit darüber, dass für einen effektiven und effizienten Dialog einzig gemischt-initiative Systeme sinnvoll sind (vgl. [Chu99, S. 2], [Chu97, S. 3]). Denn obwohl die Initiative im Grunde beim Benutzer liegen sollte, muss das System an manchen Stellen nachfragen können, um Zuverlässigkeit und Angemessenheit der Dialogantworten zu gewährleisten.

Sprachgenerierung

Bei der Sprachgenerierung, auch Antwortgenerierung oder Response Generation (RG) genannt, wird auf der Grundlage der Entscheidung des Dialogmanagers, welche Systemreaktion angemessen ist, eine für den Benutzer verständliche Systemantwort erzeugt. Der Ablauf kann dabei als reziprok zum Sprachverstehen angesehen werden (vgl. [Kel01, S. 489]). Zumeist werden im Vorhinein alle vom System möglichen Antworten als Vorlagen gespeichert und müssen dann während des Dialogs nur noch aktiviert werden.

Sprachsynthese

Die Sprachsynthese transformiert geschriebene Wörter in gesprochene Sprache und wird daher auch als TTS (Text-to-Speech) bezeichnet. Dabei unterscheidet man zwischen Konkatenation und generativer Sprachsynthese (vgl. [Pra04, S. 17f]). Bei der Konkatenation werden zuvor aufgenommene Satz- oder Wortfragmente aneinandergefügt und als kompletter Antwortsatz abgespielt. Da die einzelnen Worte zuvor von einem Menschen gesprochen und aufgezeichnet wurden, entstehen bei der Konkatenation Antworten mit hoher Tonqualität, allerdings ohne jegliche Prosodie¹¹. Dem gegenüber steht die generative Sprachsynthese, bei der die gesamte Systemantwort künstlich erzeugt wird. Der Vorteil, dass hierbei auch prosodische Merkmale bedacht werden können, wird durch eine schlechtere Tonqualität geschmälert.

¹¹ Prosodie, [gr.-(lat.)] in der antiken Metrik die Lehre von der Tonhöhe und der Quantität der Silben, Silbenmessungslehre ([Dud01]). Die Prosodie umfasst nichtverbale Elemente einer sprachlichen Äußerung wie Tonhöhe und Betonung.

2.3 Multimodale Dialogsysteme

2.3.1 Multimodalität

Die menschliche Kommunikation beinhaltet weitaus mehr Aspekte als nur die gesprochene Sprache. Beispielsweise werden Mimik, Gestik und Körperhaltung von Menschen ebenso dazu verwendet, ihren Gemütszustand oder ihre Meinung auszudrücken, wie das gesprochene Wort selbst. Häufig ist die korrekte Interpretation einer Äußerung (im Sinne dessen, was der Sprecher mitteilen wollte) erst mit Hilfe zusätzlicher Informationen möglich, die z.B. durch die Beobachtung der Handhaltung oder des Gesichtsausdrucks des Sprechers gewonnen werden können. Einen derartigen Dialog bezeichnet man als multimodal, da er mehrere Modalitäten beansprucht. Mit Modalität verweist man hier auf die menschlichen Sinne, die dazu dienen, eingehende Information zu erfassen und zu verarbeiten. So verfügt ein Mensch u.a. über visuelle und akustische Wahrnehmungsfähigkeiten.

Einen solchermaßen „natürlichen“ Dialog zwischen zwei Menschen mit einem Computer nachzustellen, ist zurzeit noch nicht möglich. Dennoch gibt es bereits erfolgreich funktionierende Dialogsysteme, die mehrere Ein- und Ausgabemodalitäten miteinander kombinieren. Beispielhaft soll hier das Projekt SmartKom¹² genannt werden, bei dem Sprache, Grafik, Gestik und Mimik kombiniert werden.

2.3.2 Vorteile eines multimodalen Dialogsystems

Ein Dialog zwischen Mensch und Computer gewinnt bereits deutlich an Natürlichkeit, sobald mehrere Modalitäten verwendet werden können. Da unterschiedliche Benutzer unterschiedliche Präferenzen im Hinblick auf die Modalitätswahl haben können, ist es darüber hinaus von Vorteil, wenn das System dem Benutzer die Wahl über die Modalität lässt. Neben diesen Vorzügen eines multimodal bedienbaren Systems

¹² SmartKom, deutsches Projekt des DFKI (mit über zehn Industriepartnern) mit dem Ziel, Konzepte für die Entwicklung völlig neuartiger Formen der Mensch-Technik-Interaktion zu erproben. Beispielsweise stellt das Teilprojekt SmartKom-Public für den Benutzer eine öffentliche Kommunikationszelle dar, an der z.B. Kinokarten reserviert oder Informationen über das Wetter gefunden werden können (vgl. [Sma04]).

nennen [Gib00] als wichtigen Pluspunkt die Synergieeffekte durch multimodale Bedienung.

„Interfaces can benefit from modality synergy on both the input and output sides of the system. On the input side, interpreting input which is conveyed redundantly and/or complementary in several modalities can increase interpretation accuracy, e.g. combining speech recognition and lipreading in noisy environments.“ [Gib00, Kap. Audio-visual and multi-modal speech-based systems]

Kellner nennt diesen Punkt „Effizienzsteigerung“. Nicht nur, dass der Benutzer weniger Zeit für die Kommunikation mit dem System benötigt, darüber hinaus ist der Dialog auch häufiger erfolgreich. Zudem spricht er von „erhöhter Benutzerfreundlichkeit“ (vgl. [Kel01, S. 489]). Diese beruht unter anderem darauf, dass unterschiedliche Sinneskanäle benutzt werden und somit die Beanspruchung des Benutzers sinkt (vgl. auch [Hed00, S. 206]).

2.3.3 Anforderungen an multimodale Dialogsysteme

In der Literatur lassen sich zahlreiche Richtlinien für die Dialogmodellierung finden. Viele stellen eine einfache Erweiterung der bekannten „Usability Guidelines“ für Benutzerschnittstellen von Jacob Nielsen oder Ben Shneiderman dar (vgl. hierzu [Nie93, S. 115ff.], [Shn02, S. 73ff.]). An dieser Stelle sollen die Richtlinien von Dybkjaer et al. näher erläutert werden, da sie sich konkret mit sprachbedienbaren und multimodalen Dialogen befassen (vgl. [Dyb04, S. 45f.]). Sie beinhalten:

- Input recognition accuracy
- Naturalness of user speech
- Output voice quality
- Output phrasing adequacy
- Feedback adequacy
- Adequacy of dialogue initiative
- Naturalness of the dialogue structure

- Sufficiency of task and domain coverage
- Sufficiency of reasoning capabilities
- Sufficiency of interaction guidance
- Error handling adequacy
- Sufficiency of adaptation to user differences
- Modality appropriateness

Zunächst einmal ist ein guter Spracherkenner notwendig (*input recognition accuracy*). Außerdem sollte der Benutzer so natürlich wie möglich mit dem Dialogsystem sprechen können (*naturalness of user speech*). Das Vokabular ist hier die entscheidende Komponente. Die Qualität der von der Sprachsynthese gelieferten Systemausgaben sollte ebenfalls möglichst hoch sein (*output voice quality*). Dies bedeutet insbesondere eine klare und verständliche Aussprache, aber auch möglichst natürliche Betonung und Sprechgeschwindigkeit. Dabei sollten die Systemausgaben korrekt, relevant und ausreichend informativ sein (*output phrasing adequacy*).

Der Benutzer sollte Rückmeldungen über die Vorgänge des Systems erhalten, um sicher sein zu können, dass seine Eingaben korrekt verstanden und verarbeitet wurden (*feedback adequacy*). In Abhängigkeit von der Aufgabe, den Erfahrungen des Benutzers und dem Aufgabenumfeld muss die passende Dialoginitiative gewählt werden (*adequacy of dialogue initiative*)(vgl. hierzu auch Kapitel 2.2 unter „Dialogsteuerung“). Dabei sollte der gesamte Aufbau des Dialogs möglichst nah an den Erwartungen des Benutzers liegen (*naturalness of the dialogue structure*). Auch im Hinblick auf die Aufgaben, die mit diesem System bewältigt werden können, und die Informationen, die es ihnen zur Verfügung stellen kann, haben die Benutzer bestimmte Erwartungen. Diese sollten ebenfalls so weit wie möglich erfüllt werden (*sufficiency of task and domain coverage*).

Dem System müssen Informationen über kausale Zusammenhänge vorliegen, damit es aus dem Benutzerverhalten die richtigen Schlüsse ziehen kann (*sufficiency of reasoning capabilities*). Sollte der Benutzer mehrfach Probleme bei der Bewältigung seiner Aufgaben haben oder beispielsweise häufiger nicht verstanden werden, müssen ihm angemessene Lösungsmöglichkeiten im Sinne einer Hilfe-Funktion vorgeschlagen werden

(*sufficiency of interaction guidance*). Treten Fehler auf, sollten diese erstens erkannt und zweitens angemessen behoben werden (*error handling adequacy*).

Das System sollte in dem Sinne adaptiv sein, dass es sich an unterschiedliche Benutzer anpassen kann (*sufficiency of adaptation to user differences*). Das bekannteste Beispiel hierfür sind Anfänger- bzw. Expertenversionen. Zudem sollte(n) je nach Situation die passende(n) Modalität(en) ausgewählt werden (*modality appropriateness*).

2.3.4 Multimodale Dialogsysteme im Kraftfahrzeug

Das automobile Umfeld erfordert noch eher multimodale Benutzerschnittstellen als beispielsweise die Mensch-Computer-Kommunikation am heimischen Rechner. Während einer Autofahrt ist insbesondere der visuelle Sinneskanal des Fahrers durch das Fahren selbst belastet. Werden nun weitere Informationen wie z.B. die Routeninformationen des Navigationssystems ebenfalls über eine visuelle Ausgabe dargestellt, birgt dies ein nicht zu unterschätzendes Gefahrenpotenzial. Wenn der Fahrer in einer heiklen Situation den Blick vom Verkehrsgeschehen abwendet, kann dies schlimme Folgen haben. Hedicke spricht hier auch von der „erhöhten Gefahr der Wahrnehmungsüberforderung aufgrund von Informationstransformation in eine visuelle Darstellung“ [Hed00, S. 203]. Demnach kann es passieren, dass der Fahrer die Information aufgrund der gewählten Modalität nicht aufnehmen kann. Multimodale Dialogsysteme bieten hier die Möglichkeit einer situationsgerechten und sicheren Kommunikation.

Die Tatsache, dass allein sprachliche Bedienung im Kraftfahrzeug nicht ausreicht, liegt in der Modalität selbst begründet. Für erfolgreiche sprachliche Kommunikation muss der Benutzer an jeder Stelle des Dialogs sämtliche Handlungsmöglichkeiten im Gedächtnis haben. Insbesondere bei der Auswahl von Elementen aus langen Listen führt dies zu Problemen, da sich der Benutzer unmöglich an alle Elemente erinnern kann, selbst wenn sie ihm vorgelesen wurden. Auch Salmen beschreibt diesen Punkt in ihrer Arbeit über multimodale Menüausgabe im Fahrzeug:

„Nachteilig ist dagegen, dass sprachliche Interaktion auf Memorierung beruht, die stark limitiert ist. Mit zunehmendem Umfang bzw. zunehmender Komplexität der Informationen steigt daher die mentale Belastung.“ [Sal02, S. 51]

Somit scheint die multimodale Bedienung, die die Unzulänglichkeiten einzelner Modalitäten durch ihre Kombination aufhebt, die optimale Lösung darzustellen. Auch [Aky01] kommen zu diesem Ergebnis. Sie begründen die Wichtigkeit multimodaler Bedienfunktionen im Auto folgendermaßen:

„Die Verwendung mehrerer Modalitäten, wie z.B. Sprache und Gestik, kann die Bedienung komplexer Systeme intuitiver gestalten. Ebenso ist eine höhere Bedienungseffizienz zu erwarten, weil unabhängige Ressourcen zum Einsatz kommen. Durch die Adaptation der Systemfunktionen und der Informationsdarstellung an den Nutzungskontext soll zudem die Arbeitslast verringert und ein aufmerksames Fahren ermöglicht werden.“ [Aky01, S. 137f.]

3 Beschreibung des Dialogsystems

Im folgenden Kapitel wird das Testsystem¹ für den Zieleingabe-Dialog detailliert beschrieben. Dabei sollen zunächst die einzelnen Komponenten erklärt und ihr Zusammenspiel erläutert werden. Im Anschluss wird die graphische, sprachliche und/oder multimodale Interaktion des Benutzers mit dem System anhand von Screenshots und Dialogbeispielen erläutert.

3.1 Die einzelnen Komponenten und ihr Zusammenspiel

Abbildung 3.1 zeigt eine Übersicht über das Testsystem. Dass das System multimodal bedienbar ist, lässt sich auf den ersten Blick erkennen. Es fallen sofort die beiden farbig markierten Datenkreisläufe ins Auge, die die beiden Bedienmöglichkeiten des Systems darstellen. Der erste steht für die Sprachbedienung des Systems und ist in Abbildung 3.1 rot eingezeichnet. Der zweite, blau hervorgehobene Kreislauf stellt die graphische Bedienung dar. Die grünen Pfeile stehen für die Steuerungsprozesse des Dialogmanagers. Auf sie wird später noch genauer eingegangen.

¹ *Testsystem*. Gemeint ist hier die Ausgangsversion des Dialogsystems, die von der Firma Speech Experts GmbH, Regensburg entwickelt wurde.

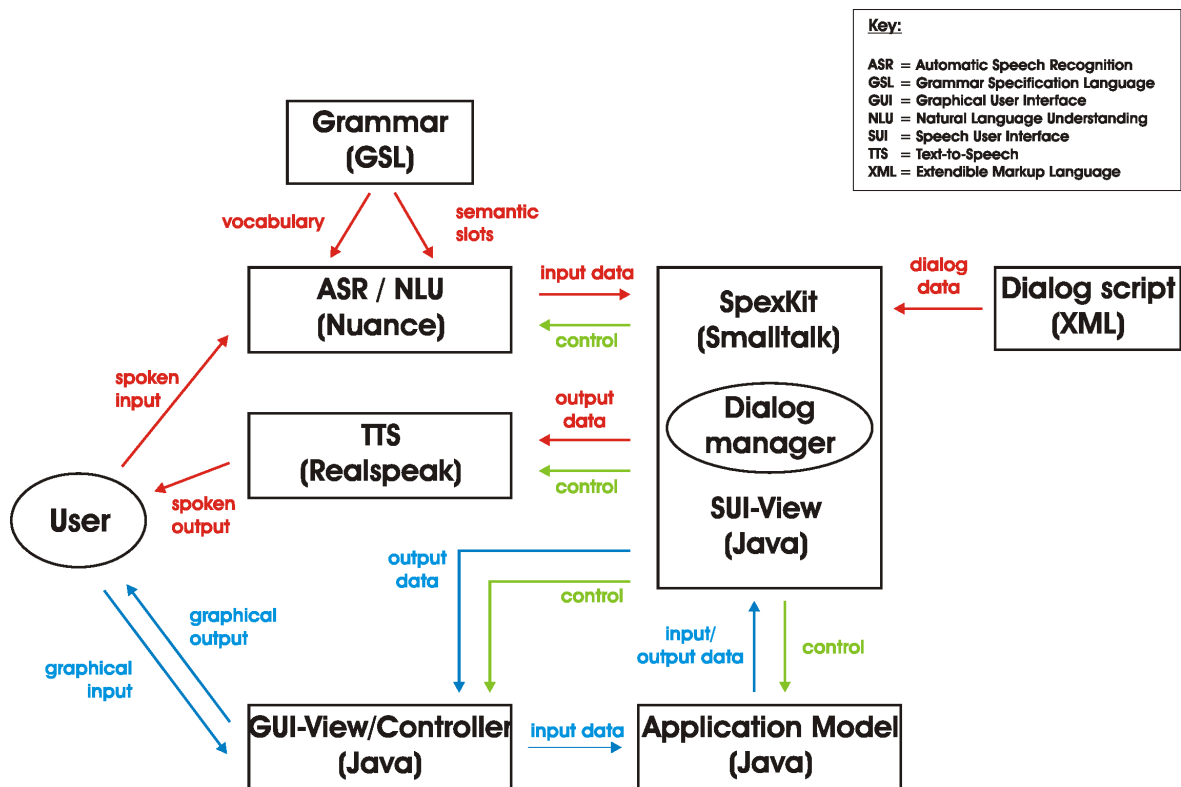


Abbildung 3.1: Das Testsystem in der Übersicht

MVC (Model-View-Controller)

Der Aufbau des gesamten Systems entspricht dem Grundgedanken eines MVC. MVC ist ein Entwurfsmuster für objektorientierte Programmiersprachen, das insbesondere in den letzten zehn Jahren an Bedeutung gewonnen hat. Dabei steht MVC für Model, View und Controller, womit die drei Komponenten innerhalb der Programmstruktur gemeint sind, die auch in Abbildung 3.2 zu sehen sind. Der Grundgedanke dabei ist, die drei Grundfunktionen eines jeden mit einem Benutzer interagierenden Systems, nämlich Eingabe, Verarbeitung und Ausgabe, strikt voneinander zu trennen. Der „Controller“ soll die Eingaben des Benutzers entgegennehmen und an das „Model“ weiterleiten. Dieses ist zuständig für die Verarbeitung der Daten und die folgende Aktionsentscheidung. Die „View“ soll daraufhin die Informationen, die sie vom „Model“ mitgeteilt bekommt, an den Benutzer weitergeben.

Die MVC-Struktur hat neben ihrer Übersichtlichkeit den Vorteil, dass sie schnell und leicht erweiterbar ist. Dies spielt insbesondere bei multimodalen Systemen eine wichtige

Rolle, da für jede neue Modalität einfach ein neuer „Controller“ und eine neue „View“ hinzugefügt werden kann.

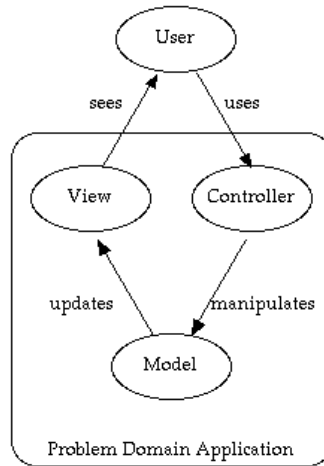


Abbildung 3.2: The model-view-controller design pattern [Bar99]

GUI-View/Controller (Java)

Eine häufig anzutreffende Vereinfachung der MVC-Struktur ist die Vereinigung von „Controller“ und „View“ zu einem User Interface (vgl. hierzu [Sun98]). So ist es auch im Testsystem bei der graphischen Benutzerschnittstelle. Der in Java programmierte „GUI-View/Controller“ nimmt die graphischen Eingaben des Benutzers über die Maus entgegen und leitet sie an das Application Model weiter. Später erhält der GUI-View/Controller vom Dialogmanager die Informationen, die der Benutzer auf dem Bildschirm sehen können soll, und stellt sie in der entsprechenden Ansicht dar (vgl. Abbildung 3.3).

Application Model (Java)

Das Application Model ist ebenfalls in Java programmiert. Es verfügt über alle graphisch eingegebenen Daten, die es vom GUI-View/Controller erhalten hat und steuert den Dialog, so lange dieser ausschließlich über die Modalität „GUI“ geführt wird. Dabei werden die Ein- und Ausgabe-Daten vom Application Model an den Dialogmanager weitergeleitet. Von dort werden dann die Informationen, die der Benutzer auf dem Bildschirm sehen können soll, an den GUI-View/Controller gesendet. Dieser „Umweg“ über den Dialogmanager wäre bei einem rein graphisch geführten Dialog nicht nötig, da das Application Model selbstständig in der Lage ist, Aktionsentscheidungen zu treffen und an den GUI-View/Controller zurückzuleiten. Allerdings soll der Benutzer dieses Systems

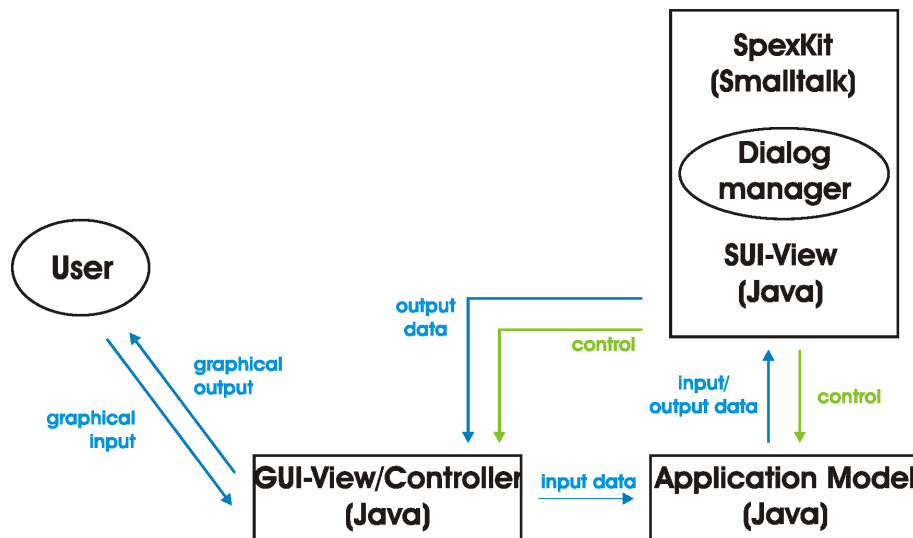


Abbildung 3.3: Die graphische Bedienung über GUI-View/Controller und Application Model

nicht nur in der Lage sein, *entweder* über die GUI *oder* über Sprache mit dem System zu interagieren, sondern er soll auch *zu jeder Zeit zwischen den Modalitäten wechseln* können. Daher ist es notwendig, dass alle Informationen aus dem graphischen Dialog zu jeder Zeit auch für den Sprachdialog zur Verfügung stehen.

ASR & NLU (Nuance)

Der hier verwendete Spracherkenner ist die Software „Nuance“, Version 8.5, von der Firma „Nuance Inc.“ in Palo Alto, USA. Dieser Spracherkenner wird hauptsächlich für server-basierte Telefondialogsysteme entwickelt und entspricht daher nicht allen Anforderungen für die Spracherkennung im Auto (insbesondere im Hinblick auf Umweltgeräusche und Ressourcenanforderungen), allerdings wird er aufgrund seiner Flexibilität für Forschungszwecke häufig auch in anderen Domänen eingesetzt.

Der Spracherkenner funktioniert sprecherunabhängig. Die angegliederte Grammatik im GSL-Format², die das gesamte Vokabular enthält, kann zur Laufzeit geladen werden. Bei dem vorliegenden System handelt es sich um ein natürlichsprachliches, das kurze Phrasen als Benutzereingabe zulässt (vgl. hierzu Kapitel 2.2 unter „Benutzereingabe“).

² GSL, *Grammar Specification Language*. Spezielles Format für Grammatiken, die von der Firma Nuance im Zusammenspiel mit dem gleichnamigen Spracherkenner verwendet werden.

Nach der sprachlichen Eingabe des Benutzers werden die vom Spracherkenner erkannten Worte bzw. Sätze semantisch interpretiert und die Ergebnisse in Slots gespeichert. Diese Slots und die Regeln für die semantische Interpretation werden ebenfalls in der GSL-Grammatik definiert. Beispielsweise gibt es einen Slot „cmd“ für das letzte vom Benutzer verwendete Kommando (z.B. „Liste zeigen“ oder „Buchstaben löschen“) und einen Slot „city“ für den Zielort. Die Inhalte dieser Slots werden als Ergebnis des Sprachverstehens an den Dialogmanager weitergegeben (vgl. Abbildung 3.4).

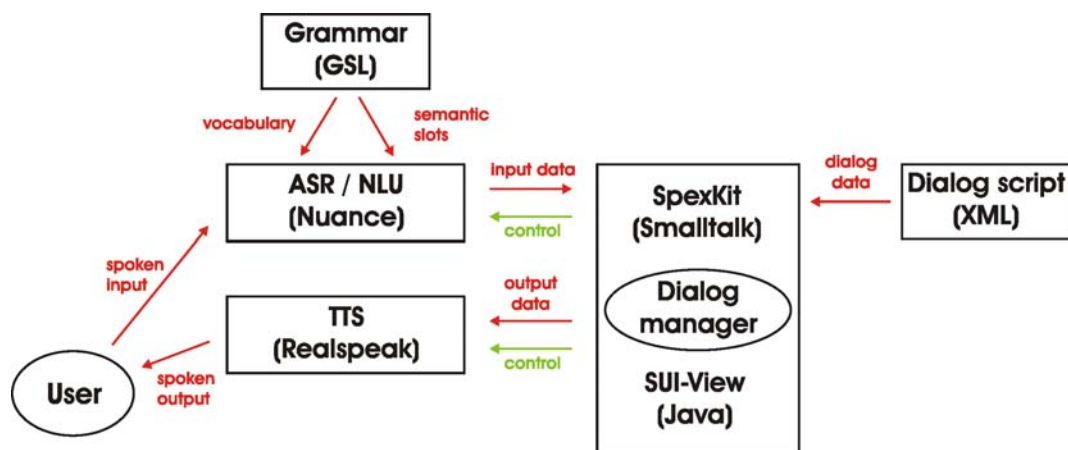


Abbildung 3.4: Die sprachliche Bedienung über ASR, NLU, Dialogskript und TTS

Dialog script (XML)

Das Dialogskript enthält die Dialogsteuerung für den gesamten Sprachdialog. Hier ist zunächst festgelegt, welche Teilgrammatik zu einem bestimmten Dialogschritt geladen wird. Somit bestimmt das Dialogskript, welche sprachlichen Benutzereingaben (d.h. Wörter/Phrasen) zu einem bestimmten Zeitpunkt möglich sind. Außerdem ist hier definiert, welche Aktion auf eine bestimmte Eingabe folgt. Beispielsweise erfolgt nach der sprachlichen Eingabe des Zielortes immer die Abfrage „Meinten Sie XY?“. Diese sprachlichen Ausgaben des Systems, die Prompts, sind ebenfalls im Dialogskript gespeichert. Dabei enthalten manche Prompts Variablen (z.B. für die Anfangsbuchstaben des Zielortes), die im Laufe des Dialogs gefüllt werden.

Das Dialogskript hat noch einige weitere Funktionen. Es legt zum einen fest, in welcher Situation welche Signaltöne abgegeben werden. Zum anderen bestimmt es, nach welcher Zeitspanne ein Timeout erfolgt. Die implementierten drei verschiedenen Timeouts werden in Kapitel 3.2.2 näher beschrieben.

Das Dialogskript wird beim Start des Testsystems vom Dialogmanager eingelesen und in einen Zustandsautomaten für die Dialogsteuerung abgebildet (vgl. Abbildung 3.4).

TTS (Realspeak)

Zur Sprachsynthese wird die Software „Realspeak“ der Firma Scansoft³ verwendet. Die künstlich erzeugte Sprache ist nur mit einer weiblichen Stimme erhältlich. Um diese möglichst natürlich klingen zu lassen, wurden einige prosodische Merkmale implementiert. Beispielsweise wird am Ende des Satzes die Stimme angehoben, wenn ein Fragezeichen folgt.⁴ Die TTS erhält im Laufe des Dialogs immer wieder Ausgabedaten vom Dialogmanager, die in Sprache umgesetzt werden (vgl. Abbildung 3.4).

Dialogmanager (SpexKit & SUI-View)

Der Dialogmanager besteht in diesem System aus zwei Komponenten. Das auf Smalltalk basierende SpexKit der Firma Speech Experts GmbH in Regensburg stellt einen zustandsbasierten Dialogmanager dar und nimmt alle Informationen von Spracherkenner und Dialogskript entgegen. Die SUI-View hingegen kommuniziert mit dem GUI-View/Controller, dem Application Model und der TTS. Beide gemeinsam steuern den Dialog. Dies ist in den Abbildungen 3.1, 3.3 und 3.4 durch die grünen Pfeile mit der Beschriftung „control“ verdeutlicht. Ein Beispiel für einen solchen Steuerungsprozess ist der Aufruf zum Starten an den Spracherkenner, sobald die Taste PTT gedrückt worden ist (vgl. hierzu auch Kapitel 3.2.2). Die Abläufe der internen Kommunikation zwischen SpexKit und SUI-View sind sehr stark ineinander verwoben, so dass sie nur gemeinsam als Dialogmanager bezeichnet werden können.

Das Dialogsystem lässt sich insgesamt als aufgabenorientiert und gemischt-initiativ bezeichnen (vgl. hierzu Kapitel 2.2, Punkt „Dialogsteuerung“). Der Dialog wird beendet, sobald das Dialogziel, die Eingabe eines Zielortes erreicht ist. Die Initiative liegt

³ Scansoft Inc., US-amerikanische Softwarefirma. Entwickelte u.a. die Text-to-Speech-Engine „Realspeak“. Für nähere Informationen zu dieser Software vgl. [Rea04].

⁴ Während des Aufbaus des Testsystems wurde festgestellt, dass die Sprachsynthese an manchen Stellen „hakt“, d.h. einzelne Silben verschluckt bzw. undeutlich ausspricht. Nach eingehender Analyse wurde jedoch erkannt, dass dies ein interner Softwarefehler von Realspeak ist. Dadurch war ein schnelles Beheben dieses Fehlers ausgeschlossen. Da allerdings größere Verständnisprobleme trotz dieses kleinen Fehlers nicht auftraten, wurde er akzeptiert und wird im weiteren Verlauf dieser Arbeit nicht weiter erwähnt werden.

während der graphischen Interaktion beim Benutzer. Der Sprachdialog ist im Normalfall systemgetrieben, da das System jeden neuen Dialogschritt mit einem Prompt bzw. einer Frage beginnt, auf die der Benutzer antwortet. Der Benutzer hat allerdings durch „Barge-In“ jederzeit die Möglichkeit, die Dialoginitiative zu ergreifen. Wenn der Benutzer allerdings drei Mal hintereinander nicht verstanden wurde, übernimmt das System wieder die Initiative und bricht den Dialog ab.

Das Dialogsystem ist weder adaptiv noch einstellbar, das heißt es passt sich nicht an den Benutzer an und lässt sich von diesem auch nicht nach persönlichen Präferenzen einstellen. Auch die Ein- und Ausgabemodalitäten sind festgelegt. Bei graphischer Eingabe gibt es auch nur eine graphische Ausgabe, bei sprachlicher Eingabe erfolgen sowohl sprachliche als auch graphische Ausgabe.

3.2 Der Dialog mit dem Testsystem

Im Folgenden werden die Interaktionsmöglichkeiten des Benutzers mit dem Testsystem näher erklärt. Nachdem zunächst anhand der einzelnen Ansichten auf die GUI-Bedienung eingegangen wird, folgen später Erläuterungen zur Sprach- und multimodalen Bedienung. Im Anschluss an diese Darstellungen werden jeweils beispielhafte Dialoge eines Benutzers mit dem Testsystem vorgestellt.

3.2.1 Bedienung über die graphische Oberfläche

Die gesamte Bedienung der GUI soll mit der Maus vollzogen werden. An einzelnen Stellen ist auch die Verwendung der Tastatur möglich. So können beispielsweise für das Navigieren in der Liste auch die Cursortasten verwendet werden. Allerdings sind nur vereinzelte Funktionen für die Tastatur implementiert. Da demnach keine einheitliche Bedienung mittels Tastatur möglich ist, wird im Folgenden nur auf die Bedienung mittels Maus eingegangen.

Hauptmenü

Abbildung 3.5 zeigt das Hauptmenü. Es hat den Titel „Navigation“. Da dieses System später einmal in ein größeres Fahrerinformationssystem integriert werden soll, das

neben dem Navigationsmenü noch weitere Menüs, beispielsweise für die MP3- oder Telefonbedienung, enthalten wird, macht diese Benennung durchaus Sinn. Der Benutzer kann so jederzeit am Titel erkennen, in welchem Menü er sich gerade befindet. Zusätzlich ist in der linken oberen Ecke ein Icon dargestellt, das auch auf allen anderen Seiten zur Zielorteingabe und Navigation zu sehen sein wird. Dieses Icon ist ein Wegweiser.



Abbildung 3.5: Testsystem: Hauptmenü

Die gesamte graphische Oberfläche wurde so programmiert, dass direktes Anklicken der einzelnen Punkte grundsätzlich möglich ist. Somit liegt hier die Simulation eines Touchscreen vor - nur die „Touch“-Funktion des Bildschirms fehlt.

Das Hauptmenü enthält nur zwei Optionen: Der Benutzer kann entweder einen Zielort eingeben (1) oder die Navigation starten (2). Letzteres macht natürlich erst dann Sinn, wenn zuvor ein Zielort eingegeben wurde. Allerdings ist im hier dargestellten System keine Zielführung realisiert, so dass bei Anklicken des Punktes „Navigation starten“ eine Art Fehlerseite geöffnet wird, die die Meldung „Kein GPS-Empfang“ enthält. Diese Seite ist in Abbildung 3.6 dargestellt. Das System schließt diese Seite selbstständig nach vier Sekunden und kehrt zum Hauptmenü zurück. Möchte der Benutzer sofort, und nicht erst nach dieser kurzen Wartezeit, zum Hauptmenü zurückkehren, kann er auf die Schaltfläche „OK“ (3) drücken.

Sobald der Benutzer im Hauptmenü die Option „Zielort eingeben“ wählt, öffnet sich die Listenansicht.



Abbildung 3.6: Testsystem: Fehlermeldung bei Klick auf „Navigation starten“

Listenansicht ohne Filter



Abbildung 3.7: Testsystem: Listenansicht ohne aktivierten Filter

Das Scrollen durch die Liste mittels Mausrad oder einer Scrollleiste ist nicht möglich. Allerdings kann sich der Benutzer innerhalb der Liste über das Kontrollfeld in der rechten unteren Ecke vor- bzw. rückwärts bewegen. Dieses Kontrollfeld ist in Abbildung 3.8 noch einmal vergrößert zu sehen.

Über die einzelnen Pfeile ▼ oder ▲ springt man einen Eintrag in der Liste nach unten bzw. oben, der Doppelpfeil ▼ oder ▲ führt eine ganze Seite, d.h. in diesem Fall sechs

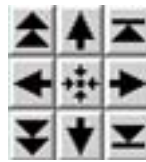





Abbildung 3.8: Testsystem: vergrößertes Kontrollfeld

Einträge, weiter bzw. zurück und über die Symbole  oder  gelangt man an das Ende bzw. den Anfang der Liste.

Der mittlere Button des Kontrollfeldes  kann zur Auswahl des markierten Eintrags benutzt werden. Die gleiche Funktion erfüllt der Button „OK“ (4) am rechten Bildrand. Eine andere Möglichkeit wäre das direkte Anklicken des gewünschten Eintrags. Die beiden Pfeile nach links  bzw. rechts  haben in dieser Ansicht keine Funktion. In der Filteransicht, in der eine Tastatur eingeblendet ist, können sie dann zum Navigieren nach rechts bzw. links auf der Tastatur verwendet werden.

Direkt über den Listeneinträgen wird dem Benutzer angezeigt, wie viele Einträge die aktuelle Liste enthält (5). Die vollständige Liste, die zu Beginn jedes Dialogs geladen wird, besteht aus 1328 Einträgen. Dabei ist die Anzeige dynamisch, d.h. sie passt sich sofort an, wenn der Benutzer über die Anfangsbuchstaben eine Einschränkung vorgenommen hat (vgl. auch Abbildung 3.11).

Am unteren Rand der Bedienoberfläche sind noch zwei weitere Schaltflächen eingeblendet. Über „Abbruch“ (6) gelangt der Benutzer zu jeder Zeit wieder zurück ins Hauptmenü. Die vorgenommenen Eingaben werden dabei gelöscht. Die Schaltfläche „Filter“ (7) führt den Benutzer in die so genannte Filteransicht (vgl. Abbildung 3.9).

Filteransicht

In der Filteransicht ist eine QWERTZ-Tastatur⁵ eingeblendet, mit der Buchstaben eingegeben werden können. Dies ist durchaus sinnvoll, da das Blättern durch die Liste bereits bei mehr als 100 Einträgen sehr lange dauert. Bei der hier implementierten Liste mit über 1300 Einträgen wäre beispielsweise das Blättern zum Buchstaben „M“ nur mit mehr als 100 Klicks möglich. Um dem Benutzer dies zu ersparen, können in der

⁵ QWERTZ-Tastatur: Die Anordnung der Buchstaben entspricht der einer Computer-Tastatur. Da dabei die ersten sechs Buchstaben (von oben links gelesen) Q, W, E, R, T und Z sind, wird diese Anordnung meist nach dieser Reihenfolge benannt.

Filteransicht die Anfangsbuchstaben des gewünschten Zielortes eingegeben werden. Die einzelnen Buchstaben können, so wie die Listeneinträge in der Listenansicht, über die simulierte Touch-Tastatur direkt angeklickt werden. Alternativ ist auch ein Navigieren über das Kontrollfeld möglich. Die ausgewählten Buchstaben erscheinen in der Filterzeile (8) über der Tastatur.



Abbildung 3.9: Testsystem: Filteransicht

Die Abbildung 3.10 zeigt einen solchen Filter nach der Eingabe. Dabei bleibt es dem Benutzer selbst überlassen, wie viele Buchstaben er eingeben will. Meist reicht allerdings die Eingabe von nur drei Buchstaben, um eine hinreichend gefilterte Liste zu erhalten.



Abbildung 3.10: Testsystem: Filteransicht mit eingegebenem Filter

Über den Button mit der Beschriftung „C“ (9), was für „Clear“ steht, kann der zuletzt eingegebene Buchstabe gelöscht werden. Das Löschen aller bisher eingegebenen Buchstaben ist über die Schaltfläche „Löschen“ (10) möglich, die nur in der Filteransicht erscheint. Die Schaltfläche „Liste>“ (11) funktioniert analog zur Schaltfläche „Filter>“ in der Listenansicht. Durch sie wechselt der Benutzer wieder zurück zur Listenansicht, wobei die Liste anhand der eingegebenen Buchstaben gefiltert wird. Eine auf diese Art und Weise gefilterte Liste ist in Abbildung 3.11 zu sehen.

Listenansicht bei aktiviertem Filter



Abbildung 3.11: Testsystem: Listenansicht mit gefilterter Liste

Der momentan eingegebene Filter wird nun in der Titelzeile mit angezeigt (12). Gut zu erkennen ist auch, dass sich die Anzeige der Anzahl der Listeneinträge angepasst hat (13). In diesem Fall wurde der Filter „Hil“ eingegeben, wodurch die Liste auf fünf Einträge eingegrenzt wurde.

Es kann passieren, dass in der Liste kein Eintrag mit den eingegebenen Anfangsbuchstaben vorhanden ist, beispielsweise wenn sich der Benutzer „vertippt“ hat oder wenn er nach einem Ort sucht, der z.B. aufgrund seiner Größe nicht in der Liste aufgeführt ist. In diesem Fall erscheint die Ansicht „Liste leer“ (vgl. Abbildung 3.12).

Hat der Benutzer einen Eintrag aus der Liste (wie sie z.B. in Abbildung 3.11 dargestellt ist) ausgewählt, wechselt das System wieder zum Hauptmenü. Dort erscheint die



Abbildung 3.12: Testsystem: Listenansicht bei leerer Liste

Anzeige des ausgewählten Zielortes, z.B. „Aktueller Zielort: Hildesheim“ (vgl. Abbildung 3.13).

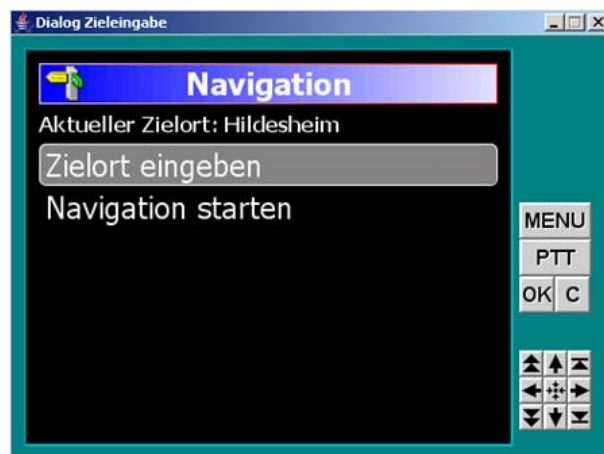


Abbildung 3.13: Testsystem: Hauptmenü mit ausgewähltem Zielort

Obwohl nun ein Zielort eingegeben wurde, springt die Markierung im Hauptmenü nicht zu „Navigation starten“, sondern bleibt auf „Zielort eingeben“ stehen.

Durch die in diesem Kapitel dargestellten Screenshots sollte deutlich geworden sein, dass es im Grunde nur drei unterschiedliche GUI-Ansichten gibt: Hauptmenü, Listenansicht und Filteransicht. Dabei können allerdings alle drei mit unterschiedlichen Eingaben versehen werden (z.B. Filteransicht mit und ohne eingegebene Buchstaben).

Nachfolgend wird mit dem Beispieldialog A eine erfolgreiche Interaktion eines Benutzers mit dem System vorgestellt. Dabei erfolgt die Bedienung hier nur über die graphische Oberfläche (GUI). Weitere Beispieldialoge zur Verdeutlichung der Sprach- und multimodalen Bedienung folgen im weiteren Verlauf des Kapitels.

Beispieldialog A

Bedienung über die graphische Oberfläche (GUI).

Der Benutzer möchte nach Stuttgart.

Tabelle 3.1: Beispieldialog A

Benutzer	Testsystem
klickt auf „Zielort eingeben“.	
	öffnet die Listenansicht.
klickt auf die Schaltfläche „Filter>“.	
	öffnet die Filteransicht.
klickt nacheinander auf die Buchstaben S, T und U.	
	zeigt in der Filterzeile (über der Tastatur) die eingegebenen Buchstaben an.
klickt auf die Schaltfläche „Liste>“.	
	wechselt zurück zur Listenansicht und schränkt die Liste anhand des eingegebenen Filters ein. Es erscheinen nur noch 2 Einträge: „Stuttgart“ und „Stuhr“.
klickt auf den Eintrag „Stuttgart“.	
	wählt Stuttgart als Zielort aus und wechselt zurück zum Hauptmenü. Dort erscheint die Anzeige „Aktueller Zielort: Stuttgart“.

3.2.2 Sprachbedienung

Möchte der Benutzer das Testsystem mittels Sprache bedienen, so muss er zunächst die Spracherkennung starten. Dazu dient der Button mit der Beschriftung „PTT“ (Push-to-Talk) (vgl. Abbildung 3.14).

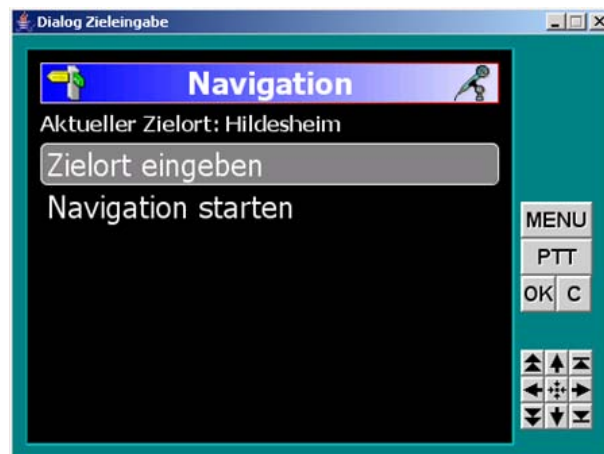


Abbildung 3.14: Testsystem: Hauptmenü mit aktivierter Spracherkennung

Über den GUI-View/Controller und das Application Model gelangt die Information, dass die Taste „PTT“ betätigt wurde, an den Dialogmanager. Diesem wird ebenso der aktuelle Dialogzustand mitgeteilt, d.h. in welcher Ansicht sich der Benutzer momentan befindet und welche Eingaben bisher getätigt wurden. Daraufhin führt der Dialogmanager die folgenden Aktionen durch: Zunächst wird der TTS der entsprechende Prompt mitgeteilt, der in ein Audiosignal umzuwandeln ist. Nachdem die Sprachausgabe beendet ist, wird vom SpexKit aus ein Tonsignal abgegeben („Beep“). Nun gibt der Dialogmanager dem Spracherkenner die Anweisung zu starten. Das kann nicht früher geschehen, da sonst der vom SpexKit abgegebene Ton mit erkannt werden würde. Ist die Spracherkennung geöffnet, geht die Anordnung, das Mikrofon-Symbol einzublenden, an den GUI-View/Controller. Beide, Ton und Symbol, sollen dem Benutzer mitteilen, dass die Spracherkennung nun aktiviert ist (Symbol-Einblendung vgl. Abbildung 3.14).

Das Abgeben des Tons sollte im Grunde gleichzeitig mit dem Einblenden des Symbols passieren, zumal es für den Benutzer die gleiche Information enthält. Da jedoch zwischen diesen beiden Schritten der Spracherkenner gestartet werden muss, gibt es hier leider eine zeitliche Verzögerung. Diese beträgt zwar nur einige Millisekunden, ist jedoch für den

Menschen bemerkbar. Das Problem dabei ist nicht, dass das Symbol zu spät eingeblendet, sondern dass der Ton zu früh abgegeben wird, nämlich wenige Millisekunden bevor der Spracherkenner geöffnet ist. Dies könnte dazu führen, dass diejenigen Benutzer, die nur auf den Ton achten, zu früh anfangen zu sprechen.

Während des laufenden Dialogs muss die Spracherkennung nicht immer wieder vom Benutzer durch einen Klick auf die „PTT“-Taste aktiviert werden. Stattdessen läuft dieser Prozess automatisch ab. Sobald die TTS einen Prompt fertig „gesprochen“ hat, folgt das Tonsignal und die Spracherkennung öffnet sich.

Nach dem Tonsignal hat der Benutzer vier Sekunden Zeit, um anzufangen zu sprechen. Sagt er nichts, geht der Dialog in den Zustand „pausierend“ über (Initial Timeout). Alle bisher eingegebenen Daten bleiben erhalten. Möchte der Benutzer den Dialog fortsetzen, muss er nur erneut die Taste „PTT“ betätigen.

Für eine Äußerung stehen dem Benutzer ab dem Zeitpunkt, zu dem er anfängt zu sprechen, zehn Sekunden zur Verfügung (Maximum Timeout). Danach schließt sich die Spracherkennung automatisch und verarbeitet die gesprochenen zehn Sekunden. Ist die Äußerung des Benutzers kürzer, wird auch dies vom System registriert. Nach einer Pause von einer halben Sekunde geht der Spracherkenner davon aus, dass die Äußerung abgeschlossen ist (Final Timeout) und beginnt mit der Verarbeitung des Gesagten. Dieses Verhalten hat einerseits den Vorteil, dass der Benutzer nicht immer die gesamten zehn Sekunden abwarten muss, bis seine Äußerung verarbeitet wird. Auf der anderen Seite bedeutet es aber auch, dass der Benutzer innerhalb seiner Äußerung keine längeren Pausen machen darf.

Der Benutzer hat jederzeit die Möglichkeit eines Barge-In. Das bedeutet, dass er die Sprachausgaben des Systems unterbrechen kann, um selber etwas zu sagen. Auch dies funktioniert über die Taste „PTT“. Wird sie gedrückt, während gerade eine Sprachausgabe des Systems erfolgt, wird diese sofort abgebrochen. Es folgen unmittelbar das Tonsignal sowie das Mikrofon-Symbol zur Verdeutlichung, dass die Spracherkennung nun geöffnet ist.

Spricht der Benutzer undeutlich, zu laut oder zu leise, kann der Spracherkenner der Äußerung keine Worte bzw. keinen Inhalt entnehmen. In diesem Fall antwortet das Testsystem mit der Ausgabe „wie bitte?“ und der Wiederholung des letzten Prompts, sofern kein zweiter für genau diesen Fall definiert wurde. Zusätzlich ertönt vor dem

„wie bitte?“ ein kurzer hoher Signalton, der auf die Fehlerkennung aufmerksam machen soll.

Während des Sprachdialogs erfolgt nicht nur eine sprachliche Ausgabe an den Benutzer, sondern auch eine graphische. Diese entspricht der graphischen Ausgabe während eines graphisch geführten Dialogs, d.h. die in Kapitel 3.2.1 vorgestellten Ansichten werden auch hier angezeigt. Zum einen hat der Benutzer dadurch auch eine visuelle Kontrollmöglichkeit seiner Eingaben, zum anderen ist dies auch die Voraussetzung dafür, dass der Benutzer jederzeit die Modalität wechseln kann.

Das Buchstabieren funktioniert bei der Spracheingabe nicht über die Buchstaben selber, sondern über das deutsche Buchstabieralphabet (vgl. Anhang A.1). Dieses besteht überwiegend aus zweisilbigen Eigennamen, die für den Spracherkenner phonetisch eindeutig zu unterscheiden sind. Zum Buchstabieren kann also nicht „A B C“, sondern muss „Anton Berta Cäsar“ gesagt werden.

Bei der Spracheingabe der Anfangsbuchstaben hat der Benutzer die Möglichkeit, mehrere Buchstaben auf einmal einzugeben. Nach dem Prompt „Bitte buchstabieren Sie“ können beliebig viele Buchstaben auf einmal eingegeben werden, sofern die Pausen zwischen den Eigennamen nicht zu lang sind. Sagt der Benutzer beispielsweise „Heinrich Ida Ludwig“, erscheinen diese drei Buchstaben in der Filterzeile und es erfolgt die Ausgabe „H I L und weiter?“. Wurde einer der eingegebenen Buchstaben nicht richtig erkannt, muss dem Benutzer die Möglichkeit gegeben werden, seine Eingabe sofort zu berichtigen. Daher wird bei dem Kommando „Korrigieren“ nicht nur der letzte Buchstabe, sondern der zuletzt eingegebene Buchstabenblock gelöscht. Diese Funktionalität ist zwar nicht konsistent zur Taste „C“ bei der GUI-Bedienung, hier jedoch trotzdem notwendig, da der Benutzer Fehler bei der Spracherkennung sofort, d.h. mit nur einem weiteren Dialogschritt, korrigieren können sollte.

In Anhang B sind die Dialogabläufe des Testsystems bei Sprachbedienung noch einmal detailliert dargestellt.

Der nun folgende Beispieldialog B beruht ausschließlich auf sprachlicher Bedienung. Dabei sollen einige typische Anfängerfehler demonstriert werden, die insbesondere dann auftreten können, wenn der Benutzer mit dem System wenig vertraut ist.

Beispieldialog B**Bedienung mit Sprache.****Der Benutzer möchte nach Hannover.**

Tabelle 3.2: Beispieldialog B

Benutzer	Testsystem
klickt auf „PTT“.	
	„Navigation.“ <i>Es folgen das Tonsignal und das Mikrofon-Symbol im Display als Zeichen, dass die Spracherkennung aktiviert wurde.</i>
„Zielort eingeben.“	
	<i>öffnet die Listenansicht. Die Liste enthält 1328 Einträge.</i> „Zielort?“
<i>wartet nicht auf Ton + Symbol</i> „Filtereingabe.“	
	<i>konnte nichts verstehen, da der Benutzer zu früh angefangen hatte zu sprechen.</i> „Wie bitte? Bitte nennen Sie den Zielort oder sagen Sie Filter, um die Liste einzuschränken.“ <i>Es folgen Ton + Symbol.</i>
„Filter einstellen.“	
	<i>öffnet die Filteransicht.</i> „Bitte buchstabieren Sie.“ <i>Es folgen Ton + Symbol.</i>
<i>hat vergessen, dass man über Eigennamen buchstabieren muss</i> „H A N.“	

Tabelle 3.2: Beispieldialog B [Fortsetzung]

Benutzer	Testsystem
	<i>konnte nichts verstehen, da der Benutzer nicht über Eigennamen buchstabiert hat. „Wie bitte? Bitte sprechen Sie die Anfangsbuchstaben des Zielortes.“ Es folgen Ton + Symbol.</i>
„Heinrich.“	
	<i>„H und weiter?“ Es folgen Ton + Symbol.</i>
„Anton Nordpol.“	
	<i>„A N und weiter?“ Es folgen Ton + Symbol.</i>
„Liste anzeigen.“	
	<i>wechselt zurück zur Listenansicht und schränkt die Liste anhand des eingegebenen Filters ein. Es erscheinen nur noch 3 Einträge: „Hanau“, „Hannoversch Münden“ und „Hannover“. „Zielort?“ Es folgen Ton + Symbol.</i>
„Ich möchte nach Hannover.“	
	<i>konnte nichts verstehen, da diese Wortkombination nicht in der Grammatik hinterlegt ist. „Wie bitte? Bitte nennen Sie den Zielort oder sagen Sie Filter, um die Liste einzuschränken.“ Es folgen Ton + Symbol.</i>
„Nach Hannover möchte ich!“	

Tabelle 3.2: Beispieldialog B [Fortsetzung]

Benutzer	Testsystem
	<i>konnte nichts verstehen, da diese Wortkombination nicht in der Grammatik hinterlegt ist.</i> „Wie bitte? Bitte nennen Sie den Zielort oder sagen Sie Filter, um die Liste einzuschränken.“ <i>Es folgen Ton + Symbol.</i>
<i>schreit</i> „HANNOVER!“	
	<i>konnte nichts verstehen, da das Audiosignal übersteuert wurde.</i> „Sie wurden wiederholt nicht richtig verstanden. Abbruch.“ <i>wechselt zurück zum Hauptmenü. Die eingegebenen Daten werden gelöscht.</i>

Der Beispieldialog C zeigt den Sprachdialog eines geübten Benutzers mit dem System.

Beispieldialog C

Bedienung mit Sprache.

Der Benutzer möchte nach Hildesheim.

Tabelle 3.3: Beispieldialog C

Benutzer	Testsystem
klickt auf „PTT“.	
	„Navigation.“ <i>Es folgen Ton + Symbol.</i>
„Zielort.“	

Tabelle 3.3: Beispieldialog C [Fortsetzung]

Benutzer	Testsystem
	<p>öffnet die Listenansicht. Die Liste enthält 1328 Einträge.</p> <p>„Zielort?“</p> <p>Es folgen Ton + Symbol.</p>
„Filter.“	
	<p>öffnet die Filteransicht.</p> <p>„Bitte buchstabieren Sie.“</p> <p>Es folgen Ton + Symbol.</p>
„Heinrich Ida Ludwig.“	
	<p>zeigt in der Filterzeile die eingegebenen Buchstaben an.</p> <p>„H I L und weiter?“</p> <p>Es folgen Ton + Symbol.</p>
„Liste zeigen.“	
	<p>wechselt zurück zur Listenansicht und schränkt die Liste anhand des eingegebenen Filters ein. Es erscheinen nur noch 5 Einträge.</p> <p>„Zielort?“</p> <p>Es folgen Ton + Symbol.</p>
„Hildesheim.“	
	<p>Hildesheim wird graphisch hervorgehoben.</p> <p>„Meinten Sie Hildesheim?“</p> <p>Es folgen Ton + Symbol.</p>
„Ja.“	
	<p>wählt Hildesheim als Zielort aus und wechselt zurück zum Hauptmenü. Dort erscheint die Anzeige „Aktueller Zielort: Hildesheim“.</p>

3.2.3 Multimodale Bedienung

Von multimodaler Bedienung wird gesprochen, wenn der Benutzer mehr als eine Modalität zur Interaktion mit dem System verwendet. Dabei soll in dieser Arbeit zwischen einmaligem Modalitätswechsel und multimodaler Eingabe unterschieden werden. Die Ausgabe wird hier nicht weiter betrachtet, da sie bereits bei einem sprachlich geführten Dialog multimodal ist (vgl. hierzu Kapitel 3.2.2). Ein einziger Modalitätswechsel kann z.B. dann vorliegen, wenn bei der gewählten Anfangsmodalität Schwierigkeiten mit der Bedienung auftreten. Dabei ist es nicht von Bedeutung, ob der Modalitätswechsel von der graphischen Bedienung zur sprachlichen oder umgekehrt geschieht. Dementsprechend wird von multimodaler Eingabe gesprochen, wenn mehr als ein Modalitätswechsel stattfindet. Das bedeutet insbesondere, dass zwischen den Modalitäten hin- und hergewechselt wird.

Der Dialogmanager des Testsystems hat, dank der in Kapitel 3.1 dargestellten Architektur des Systems, während des laufenden Dialogs zu jedem Zeitpunkt alle Informationen über den Dialogzustand zur Verfügung. Daher kann zu jedem Zeitpunkt die Modalität gewechselt werden.

Klickt der Benutzer während des laufenden Sprachdialogs auf die GUI, stellt dies einen Modalitätswechsel von der Sprach- zur GUI-Bedienung dar. Dabei wird der Sprachdialog abgebrochen. Befindet sich der Dialog gerade mitten in einer Systemausgabe der TTS, wird diese sofort beendet. Auch die Spracherkennung wird nicht mehr geöffnet. Die graphischen Ausgaben bleiben aber ebenso wie die bisher eingegebenen Daten bestehen, so dass der Benutzer den Dialog graphisch fortführen kann. Ein Modalitätswechsel von der GUI- zur Sprachbedienung liegt dann vor, wenn der Benutzer inmitten eines graphisch geführten Dialogs die Taste „PTT“ betätigt. In diesem Fall wird, wie in Kapitel 3.2.2 beschrieben, der zu diesem Dialogzustand passende Prompt von der TTS ausgegeben und die Spracherkennung gestartet. Nun kann der Dialog sprachlich fortgesetzt werden.

Bei multimodaler Eingabe, wenn also mehr als ein Modalitätswechsel vorliegt, bleibt der soeben beschriebene Ablauf der gleiche. Es kann immer wieder von der Sprach- zur GUI-Bedienung und umgekehrt gewechselt werden, da gewährleistet ist, dass bei einem Modalitätswechsel keine Informationen verloren gehen. Häufig erreicht der Benutzer bei multimodaler Bedienung sein Ziel schneller als bei unimodaler, da er je nach Situation die passendere Modalität auswählen kann.

Der nun folgende Beispieldialog D wird multimodal geführt. Der Benutzer entschließt sich in den Situationen, in denen er mit der gewählten Modalität nicht weiter kommt, die Modalität zu wechseln. Darüber hinaus spielen natürlich auch persönliche Präferenzen bei der Modalitätswahl eine große Rolle.

Beispieldialog D

Multimodale Eingabe.

Der Benutzer möchte nach Göttingen.

Tabelle 3.4: Beispieldialog D

Benutzer	Testsystem
klickt auf „Zielort eingeben“.	
	<i>öffnet die Listenansicht. Die Liste enthält 1328 Einträge.</i>
klickt auf „PTT“.	
	„Zielort?“ <i>Es folgen Ton + Symbol.</i>
„Göttingen.“	
	<i>Der Eintrag „Hattingen“ wird graphisch hervorgehoben.</i> „Meinten Sie Hattingen?“ <i>Es folgen Ton + Symbol.</i>
klickt auf „Filter>“.	
	<i>öffnet die Filteransicht.</i>
klickt nacheinander auf die Buchstaben G, Ö und T.	
	<i>zeigt in der Filterzeile die eingegebenen Buchstaben an.</i>
klickt auf „PTT“.	
	„G Ö T und weiter?“ <i>Es folgen Ton + Symbol.</i>
„Liste zeigen.“	

Tabelle 3.4: Beispieldialog D [Fortsetzung]

Benutzer	Testsystem
	<i>wechselt zurück zur Listenansicht und schränkt die Liste anhand des eingegebenen Filters ein. Es erscheint nur noch 1 Eintrag.</i> „Ziel-“ <i>Die Systemausgabe wird durch den nächsten Klick abgebrochen.</i>
klickt auf Göttingen.	
	<i>wählt Göttingen als Zielort aus und wechselt zurück zum Hauptmenü. Dort erscheint die Anzeige „Aktueller Zielort: Göttingen“.</i>

4 Entwicklung der Variante

In diesem Kapitel wird die Entwicklung der Variante zum Testsystem erläutert. Zunächst wird die durchgeführte explorative Studie vorgestellt, um daraufhin die aus den Untersuchungsergebnissen gezogenen Schlüsse besser veranschaulichen zu können.

4.1 Durchführung der explorativen Studie

Die Motivation, eine explorative Studie durchzuführen, entstand aus dem Wunsch, erst Fehler und kritische Stellen im derzeitigen Dialog aufzudecken, bevor mit der tatsächlichen Entwicklung von Varianten begonnen wird. Daher lag hierauf auch das Hauptaugenmerk während der Durchführung der Studie: Wo liegen die Schwachstellen im Dialog? Gibt es bereits sinnvolle Verbesserungsvorschläge? Außerdem sollte, sofern möglich, gleich überprüft werden, inwiefern die Versuchspersonen Kommandos bzw. kurze Sätze verwenden und ob das Konzept „PTT“, so wie es in Kapitel 3.2.2 dargelegt wurde, verstanden wird.

4.1.1 Struktur und Aufgabenstellung

Vor der Durchführung des Tests wurde das in der GSL-Grammatik hinterlegte Buchstabieralphabet von 26 auf 77 Wörter (vgl. Anhang A.2) erweitert. Die Auswahl wurde anhand der Ergebnisse der Studie von [Rie01]¹ vorgenommen. Mit den ausgewählten 77 Wörtern lässt sich laut [Rie01] eine Abdeckung von ca. 90% erreichen.

¹ In Riepers Studie waren 29 Testpersonen aufgefordert worden, jeweils 40 Namen deutscher Städte zu buchstabieren. Das Hauptaugenmerk lag dabei auf dem verwendeten Vokabular beim Buchstabieren mit Eigennamen.

Getestet wurden acht Personen (4 männlich, 4 weiblich), die sich als Mitarbeiter der Abteilung FV/SLI, in der diese Magisterarbeit geschrieben wurde, freiwillig als Versuchspersonen zur Verfügung gestellt hatten. Ein Test dauerte im Durchschnitt ungefähr 45 Minuten. Davon wurden ca. 10 Minuten für die Begrüßung und Einführung verwendet, ca. 20 Minuten dauerte der Versuch selbst und ca. 15 Minuten brauchten die Versuchspersonen, um den Fragebogen auszufüllen und abschließend ihre Meinung kundzutun.

Der Test fand in einem gesonderten Raum statt, damit möglichst wenig Lärm von außen die Spracherkennung und die Probanden stören konnte. Alle Aufgaben waren am Laptop zu bewältigen. Dabei konnten sich die Versuchspersonen ganz auf die Aufgabenstellungen konzentrieren - sie mussten nicht nebenbei noch eine Fahraufgabe o.ä. erfüllen. Bevor die Versuchspersonen mit den Aufgaben begannen, erklärte die Versuchsleiterin kurz den Zweck und die Bedienung des Systems. Anschließend wurden beispielhaft zwei kurze Aufgaben von der Versuchsleiterin vorgeführt. Die Aufgabenstellung war fest vorgegeben und wurde von den Versuchspersonen selbst gelesen. Sie bestand aus 14 zu bewältigenden Aufgaben, von denen die ersten fünf nur über die GUI, die zweiten fünf nur mit Sprache und die letzten vier frei nach Wahl bedient werden sollten. Bei den letzten vier Aufgaben ist zu beachten, dass zum Zeitpunkt der Durchführung dieser ersten Studie noch keine multimodale Bedienung möglich war. Dieses „Feature“ wurde erst nachträglich aus Regensburg geliefert. Demnach konnte nur entweder über die GUI oder mit Sprache bedient werden, ein Modalitätswechsel inmitten einer Aufgabe war nicht möglich. Es folgen drei beispielhafte Aufgaben:

1. Bedienung mit der Maus:

Geben Sie als Zielort „Apolda“ ein. Benutzen Sie dazu die Maus und die Funktionen zum Weiter- bzw. Zurückblättern.

2. Bedienung mittels Sprache:

Geben Sie nun als Zielort „Münster“ ein. Benutzen Sie dazu ausschließlich Sprachanweisungen und versuchen Sie, ohne Buchstabieren auszukommen.

3. Bedienung nach Belieben:

Geben Sie nun so schnell wie möglich den Zielort „Bad Homburg“ ein.

Die komplette Aufgabenstellung sowie der Fragebogen zur explorativen Studie ist dem Anhang C zu entnehmen.

4.1.2 Ergebnisse

Die Ergebnisse der Studie, die nun vorgestellt werden, wurden aus dem Fragebogen, dem Versuchsprotokoll der Versuchsleiterin und dem abschließenden Gespräch mit den Probanden gewonnen. Dabei soll nach Ergebnissen, die die GUI betreffen und solchen, die den Sprachdialog betreffen, unterschieden werden. Es wird darauf verzichtet, die genaue Anzahl der Probanden zu nennen, bei denen die einzelnen Ereignisse auftraten.

GUI:

- Die Beschriftung „PTT“ wurde nicht verstanden.
- Die Beschriftung „Filter>“ wurde nicht verstanden.
- Die Taste zum Korrigieren („C“) wurde nie benutzt.
- Die Taste „OK“ wurde nur sehr selten benutzt.
- Eine Leertaste zum Eingeben von Zielorten, die aus zwei Wörtern bestehen (z.B. „Bad Hersfeld“) wurde vermisst.
- Eine Taste „Zurück“ wurde vermisst.
- Die Qwertz-Tastatur wurde als verwirrend empfunden, da eine Eingabe über die Computer-Tastatur nicht möglich war. Es wurde sehr lange nach den richtigen Buchstaben gesucht.
- Das graphische Hervorheben des angeklickten Ortes wurde vermisst.
- Nach erfolgreicher Zieleingabe wurde die graphische Markierung im Hauptmenü als verwirrend empfunden („Zielort eingeben“ bleibt markiert, obwohl nun bereits ein Zielort eingegeben wurde).
- Die Modalität GUI wurde bevorzugt benutzt.

Sprachdialog:

- Das Buchstabieren über Eigennamen funktionierte nur mit Vorlage (ausgedrucktes Buchstabieralphabet). Ohne Vorlage waren die Probanden überfordert, wenn sie neben der Sprachbedienung auch noch über einen Eigennamen zu dem jeweiligen

Buchstaben nachdenken sollten. Dies zeigte sich in immer wieder auftretenden Timeouts.

- Das Buchstabieren wurde häufig bereits in der Listenansicht versucht.
- Die Systemausgaben wurden häufig als nicht hilfreich empfunden (Beispiel: „Navigation.“).
- Die Systemausgabe „wie bitte?“ wurde als unhöflich empfunden.
- Es wurden häufig Worte benutzt, die nicht in der Grammatik hinterlegt waren.
- Es wurden von allen acht Versuchspersonen Kommandos benutzt, keine kurzen Sätze. Allerdings haben es zwei Versuchspersonen zunächst mit kurzen Sätzen versucht, sind aber schließlich zu Kommandos übergegangen, nachdem sie mehrfach nicht verstanden wurden.
- Wenn die Benutzer nicht verstanden wurden, änderten Sie nur sehr selten ihr Verhalten. Häufig waren sie der Meinung, wenn sie nur deutlicher sprechen würden, könnte das System sie auch verstehen. Dass die Fehlerkennungen häufig an der Lautstärke, mit der gesprochen wurde, lagen, oder daran, dass Worte benutzt wurden, die nicht in der Grammatik hinterlegt waren, schien ihnen nicht klar zu sein. Insbesondere fehlte ihnen das Verständnis für Fehlerkennungen, die aufgrund von ähnlichen Phonemen oder sogar Homophonen auftraten (z.B. „Göppingen“ statt „Göttingen“ oder „Aalen“ statt „Ahlen“).
- Die zeitliche Verzögerung zwischen dem Tonsignal und der Öffnung des Spracherkenners sorgte für Verwirrung. Fast allen Versuchspersonen war grundsätzlich unklar, wann sie sprechen durften und wann nicht. Der häufigste Fehler war, dass zu früh versucht wurde, zu sprechen (d.h. bevor der Spracherkenner geöffnet war).
- Insgesamt wurde der Ton jedoch als hilfreicher Feedback empfunden als das Mikrofon-Symbol. Die meisten Versuchspersonen gaben an, nie auf das Symbol, sondern immer nur auf den Ton geachtet zu haben.

Die Studie deckte auch noch einige kleinere Fehler auf. Beispielsweise gab es Wörter, die zwar in der Grammatik hinterlegt waren, allerdings vom Sprachdialog nicht weiterverarbeitet wurden. Verwendete der Benutzer ein solches Wort, blieb der Dialog einfach stehen. Genauso gab es Probleme mit dem Barge-In. Die Systemausgabe wurde zwar

abgebrochen, der Spracherkenner jedoch nicht wieder geöffnet, so dass der Benutzer keine neue Eingabe machen konnte. Mit der Aufdeckung eben solcher Fehler und den oben genannten Ergebnissen war die Studie insgesamt erfolgreich.

4.2 Weiterentwicklung des Dialogs

Aus den oben vorgestellten Ergebnissen der explorativen Studie wurden Schlussfolgerungen dahingehend gezogen, welche Änderungen für die zu entwickelnden Varianten zu machen seien. Auch diese Änderungen sollen getrennt nach GUI und Sprachdialog vorgestellt werden.

4.2.1 GUI-Variante



Abbildung 4.1: Vergleich Testsystem (links) und Variante (rechts): Listenansicht

Mikrofon-Symbol statt Beschriftung „PTT“

In Abbildung 4.1 lassen sich einige der vorgenommenen Änderungen sofort erkennen. Die Beschriftung „PTT“ wurde herausgenommen und an ihrer Stelle ein Mikrofon eingefügt. Diese bildhafte Kennzeichnung der Taste zum Starten der Spracherkennung hat den Vorteil, dass selbst Benutzer, die dieses System noch nie verwendet haben, eine Vorstellung davon haben, was bei Druck auf die Taste passiert. Somit wird die Selbsterklärungsfähigkeit des Systems verstärkt. Um den Dialog insgesamt möglichst konsistent zu gestalten, wurde als Symbol das gleiche Mikrofon verwendet wie jenes, das bei geöffneter Spracherkennung in der rechten oberen Ecke des Displays erscheint.

Beschriftung „Abbrechen“ statt „Abbruch“

Die Beschriftung „Abbruch“ wurde in „Abbrechen“ umbenannt. Den Anstoß dazu gab die Überlegung, dass in der heute am weitesten verbreiteten Software (Microsoft Windows/Office®) die Beschriftung „Abbrechen“ geläufig ist. Um den Benutzern eine Umgewöhnung zu ersparen, wurde diese Beschriftung übernommen. Darüber hinaus wurde die Schaltfläche von unten links nach unten rechts verschoben, da sie dort auch in der Windows Software ihren Platz hat.

Beschriftung „ABC“ statt „Filter>“

Die Beschriftung „Filter>“, die in der explorativen Studie häufig nicht verstanden worden war, wurde in „ABC“ umbenannt. Auch unerfahrene Benutzer können nun erkennen, dass sich hinter dieser Schaltfläche die Möglichkeit zum Buchstabieren befinden wird. Analog zur Verschiebung der Taste „Abbrechen“ wurde diese Schaltfläche von unten rechts nach unten links verschoben.

Schaltfläche „Zurück“ hinzugefügt

Eine Schaltfläche „Zurück“ wurde eingefügt. Diese hat, im Gegensatz zur Schaltfläche „Abbrechen“, über die man grundsätzlich ins Hauptmenü gelangt, die Funktion, den zuletzt ausgeführten Schritt rückgängig zu machen. Hat der Benutzer soeben von der Filter- zur Listenansicht gewechselt, kann er über die Schaltfläche „Zurück“ wieder in die Filteransicht gelangen. War hingegen die zuletzt ausgeführte Aktion des Benutzers das Eingeben von Buchstaben, wird bei Klick „Zurück“ der zuletzt eingegebene Buchstabe gelöscht. Die Schaltfläche „Zurück“ kann dabei beliebig häufig betätigt werden: die Anzahl der Schritte, die zurückgegangen wird, entspricht der Anzahl der Klicks.

Tasten „OK“ und „C“ entfernt

In der Bedienleiste am rechten Rand wurden die beiden Buttons „OK“ und „C“ entfernt. Beide wurden von den Probanden der explorativen Studie selten bis gar nicht verwendet. Da bei der gesamten Bedienung des Systems immer ein direktes Anklicken möglich ist, konnte die „OK“-Taste ohne Bedenken entfernt werden. Anstatt der Taste „C“ wurde eine Backspace-Taste implementiert, die sich, wie bei einer Computer-Tastatur, in der Filteransicht in der rechten oberen Ecke der Tastatur befindet (vgl. Abbildung 4.2).

Alphabetische Anordnung statt Qwertz-Tastatur

Die Qwertz-Tastatur wurde entfernt, da sie während der explorativen Studie nur unnötig

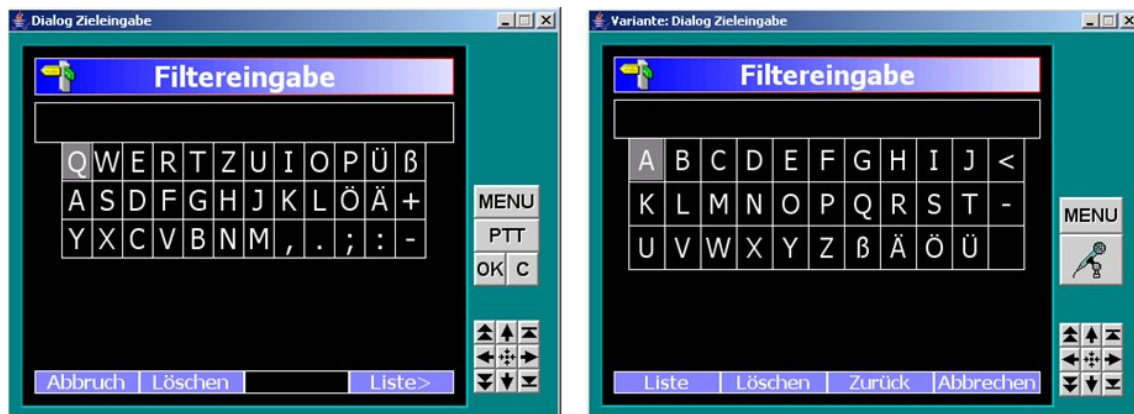


Abbildung 4.2: Vergleich Testsystem (links) und Variante (rechts): Filteransicht

für Verwirrung gesorgt hatte. Da die Eingabe der Buchstaben nicht über die Computertastatur, sondern nur mit der Maus möglich war, hatte diese Anordnung der Buchstaben zu langen „Suchzeiten“ geführt. Stattdessen wurden nun die Buchstaben in alphabetischer Reihenfolge angeordnet (vgl. Abbildung 4.2). Gleichzeitig wurden unnötige Zeichen (Komma, Semikolon, Punkt, Doppelpunkt), die bei der Eingabe eines Zielortes nicht benötigt werden, gelöscht. Hinzugefügt wurde eine Leertaste, um das Eingeben von Orten, die aus zwei Wörtern bestehen (z.B. „Groß Gerau“) zu ermöglichen.

Markierung springt auf „Navigation starten“

Nach erfolgreicher Zieleingabe blieb die Markierung bisher auf „Zielort eingeben“ stehen. Dies wurde nun geändert. Ist ein Zielort im Zielspeicher vorhanden (wie beispielsweise „Hildesheim“ in Abbildung 4.3), wird im Hauptmenü die Option „Navigation starten“ hervorgehoben.

Markierung des angeklickten Elements

Während der explorativen Studie wurde das graphische Hervorheben des angeklickten Zielortes in der Listenansicht vermisst. Das Gleiche galt für die beiden Optionen im Hauptmenü. Auch sie wurden nicht graphisch hervorgehoben; stattdessen wechselte sofort die Ansicht. Um dem Benutzer hier eine bessere Rückmeldung zu geben, wie es in Kapitel 2.3 unter dem Punkt „Anforderungen an multimodale Dialogsysteme“ mit „feedback adequacy“ bezeichnet wurde, wurde das graphische Hervorheben der angeklickten Orte bzw. Optionen ermöglicht. Hierzu wurde die jeweilige Aktion, die zuvor auf das

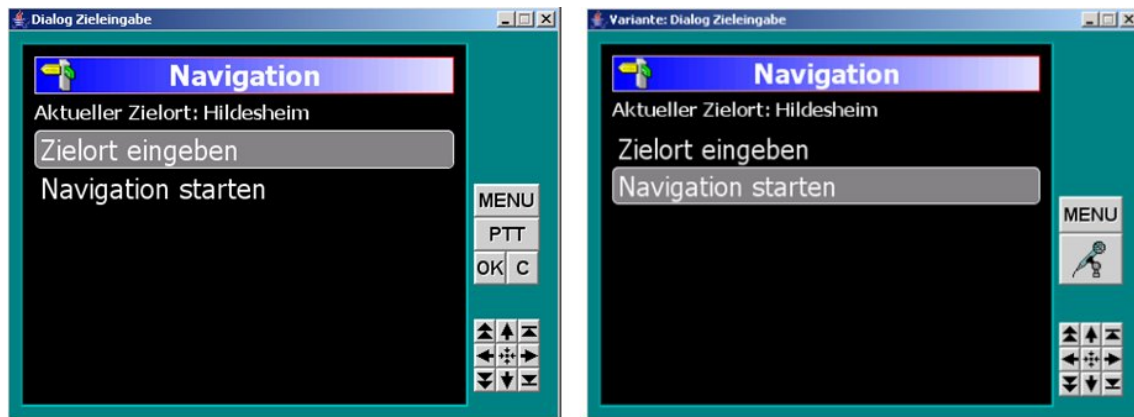


Abbildung 4.3: Vergleich Testsystem (links) und Variante (rechts): Hauptmenü

Java-Event „onmouseclicked²“ reagierte, dem Java-Event „onmousereleased³“ zugeordnet. Vorgeschaltet wurde das graphische Hervorheben auf das Java-Event „onmousepressed⁴“. Das bedeutet, dass der Ort zunächst markiert wird, sobald er angeklickt wird. Aber erst mit dem Loslassen der Maustaste wird die „wirkliche“ Aktion, z.B. die Auswahl des Zielortes, durchgeführt. Dieses Verhalten wurde für alle Ansichten gleich implementiert.

Alle Änderungen, die die GUI betreffen, wurden im Java-Code vorgenommen. Tabelle 4.1 zeigt sie noch einmal als Übersicht.

Tabelle 4.1: Übersicht GUI-Änderungen

Testsystem	Variante
Taste „PTT“	Taste mit Mikrofon-Symbol
Schaltfläche „Abbruch“	Schaltfläche „Abbrechen“
Schaltfläche „Filter>“	Schaltfläche „ABC“
—	Schaltfläche „Zurück“

² *onmouseclicked*, in Java eine Methode des Interface „MouseListener“ im Package „java.util“. „Invoked when the mouse button has been clicked (pressed and released) on a component.“ (<http://java.sun.com/j2se/1.4.2/docs/api/>, 27.12.2004)

³ *onmousereleased*, „Invoked when the mouse button has been released on a component.“ (Quelle s.o.)

⁴ *onmousepressed*, „Invoked when the mouse button has been pressed on a component.“ (Quelle s.o.)

Tabelle 4.1: Übersicht GUI-Änderungen [Fortsetzung]

Testsystem	Variante
Taste „OK“	—
Taste „C“	Backspace-Taste
Qwertz-Tastatur	Alphabetische Anordnung der Buchstaben
Komma, Semikolon, Punkt, Doppelpunkt	—
—	Leerzeichen
im Hauptmenü: Markierung bleibt auf „Zielort eingeben“ stehen	im Hauptmenü: Markierung wechselt bei eingegebenem Zielort auf „Navigation starten“
in allen Ansichten: keine Markierung bei Klick	in allen Ansichten: Markierung des angeklickten Ortes

4.2.2 Sprachdialog-Variante

Zunächst werden nun wie im vorherigen Kapitel die einzelnen Änderungen nacheinander vorgestellt. Eine Übersicht über alle Änderungen am Sprachdialog findet sich am Ende des Kapitels (Tabelle 4.6).

Buchstabieren auch in der Listenansicht

Das Buchstabieren wird nun, wie es in der explorativen Studie häufig versucht worden war, bereits in der Listenansicht zugelassen. Ebenso ist jetzt auch das Korrigieren und Löschen von Buchstaben in der Listenansicht möglich. Durch diese Änderung kommt ein wesentlicher Vorteil der Sprachbedienung - nämlich dass bei der richtigen Wortwahl nicht erst zu einer bestimmten Ansicht gewechselt werden muss, um eine Eingabe zu machen - erst wirklich zum Tragen. Wird bei der Eingabe eines Zielortes buchstabiert, verkürzt sich die Anzahl der Dialogschritte um mindestens zwei, und zwar um genau die beiden Schritte, die man vorher benötigte, um von der Listen- zur Filteransicht und wieder zurück zu wechseln.

Änderung der Prompts

Die Systemausgaben des Testsystems waren häufig als wenig hilfreich empfunden worden. Beispielsweise enthielt der Prompt „Navigation“, der im Hauptmenü ausgegeben wird, keinerlei Handlungshinweis. Daher wurden die Systemausgaben, wie in den Tabellen 4.2 bis 4.5 dargestellt, geändert bzw. hinzugefügt.

Tabelle 4.2: Prompts im Hauptmenü

Bezeichnung	Prompt im Testsystem	Prompt in der Variante
p_navi	Navigation.	Zielort eingeben oder die Zielführung starten
p_navi_2	Zielort eingeben oder Zielführung starten?	Möchten Sie den Zielort eingeben oder die Navigation starten?

Tabelle 4.3: Prompts in der Listenansicht

Bezeichnung	Prompt im Testsystem	Prompt in der Variante
p_enterCity	Zielort?	Bitte nennen Sie den Zielort.
p_enterCity_2	Bitte nennen Sie den Zielort oder sagen Sie Filter um die Liste einzuschränken.	Bitte nennen Sie den Zielort oder sagen Sie die Anfangsbuchstaben.
p_continue_list	—	Bitte sprechen Sie weitere Buchstaben oder nennen Sie den Zielort. Sie können außerdem korrigieren oder alle Buchstaben löschen.
p_confirm	Meinten Sie X?	Meinten Sie X?
p_confirm_2	—	Bitte antworten Sie mit ja oder nein.

Tabelle 4.3: Prompts in der Listenansicht [Fortsetzung]

Bezeichnung	Prompt im Testsystem	Prompt in der Variante
p_readlist	—	Die Liste wird jetzt vorgelesen. Mit ja können Sie ein Element direkt nach dem Vorlesen auswählen.
p_listend	Ende der Liste.	Ende der Liste.
p_pageend	—	Ende der Seite.
p_pageend_todo	—	Nächste Seite vorlesen?
p_pageend_todo_2	—	Soll die nächste Seite vorgelesen werden?
p_listempty	Die Liste ist leer. Abbruch oder Filter?	Die Liste ist leer.
p_listempty_2	—	Mit den Anfangsbuchstaben XYZ gibt es keine Stadt in der Datenbank.
p_listempty_todo	—	Sie können die Eingabe korrigieren oder abbrechen.

Tabelle 4.4: Prompts in der Filteransicht

Bezeichnung	Prompt im Testsystem	Prompt in der Variante
p_enterLetters	Bitte buchstabieren Sie.	Bitte buchstabieren Sie.
p_enterLetters_2	Bitte sprechen Sie die Anfangsbuchstaben des Zielortes.	Bitte sprechen Sie die Anfangsbuchstaben des Zielortes.
p_continue	und weiter?	und weiter?

Tabelle 4.4: Prompts in der Filteransicht [Fortsetzung]

Bezeichnung	Prompt im Testsystem	Prompt in der Variante
p_continue_2	Bitte sprechen Sie weitere Buchstaben oder sagen Sie Ende. Sie können außerdem korrigieren oder alle Buchstaben löschen.	Bitte sprechen Sie weitere Buchstaben oder sagen Sie Ende. Sie können außerdem korrigieren oder alle Buchstaben löschen.
p_correct	Korrektur!	Korrektur!

Tabelle 4.5: Allgemeine Prompts

Bezeichnung	Prompt im Testsystem	Prompt in der Variante
p_sorry	Wie bitte?	Wie bitte?
p_abort	Sie wurden wiederholt nicht richtig verstanden.	Sie wurden wiederholt nicht richtig verstanden.
p_cancel	Abbruch.	Abbruch.
p_waitForBeep	—	Bitte sprechen Sie erst nach dem Signalton!

Dass die Prompts durch die Änderungen zumeist länger wurden, wurde in Kauf genommen. Die Systemausgabe „wie bitte?“, die immer dann ausgegeben wurde, wenn der Spracherkenner nichts erkannt hatte, wurde so belassen, obwohl sie als unhöflich empfunden worden war. Die Gründe dafür waren, dass die Ausgabe erstens sehr kurz ist und zweitens längeren Ausgaben wie „Sie wurden nicht verstanden“ oder „Bitte wiederholen Sie Ihre Eingabe“ inhaltlich in nichts nachsteht. Wahrscheinlich regt der Prompt „wie bitte?“ sogar eher dazu an, das Gesagte noch einmal umzuformulieren, was bei dem vorliegenden System, das nur definierte Phrasen als Benutzereingabe zulässt, entscheidend für das systemseitige „Verstehen“ der Eingabe sein kann.

Erweiterung der Grammatik

Die GSL-Grammatik für den Spracherkenner wurde erweitert. Einzelne Worte wie „Feh-

ler“, „nein“ und „korrekt“ wurden ebenso ergänzt wie die kurzen Formulierungen „noch mal neu“ oder „N wie Nordpol“. Zusätzlich wurde bedacht, dass sich im Deutschen die Satzstellung ändert, sobald beispielsweise eine Zeitangabe hinzukommt. Aus „Ich möchte den Zielort eingeben“ wird dann „Jetzt möchte ich den Zielort eingeben“. Im Testsystem konnten derartige Phrasen nicht verstanden werden, da sie nicht in der Grammatik hinterlegt waren. Durch die Erweiterungen ergibt sich für den Benutzer also eine Vielzahl von neuen Formulierungsmöglichkeiten. Diese können auch deswegen von Bedeutung sein, weil in der explorativen Studie zwei Versuchspersonen zunächst versucht hatten, über kurze Phrasen (anstatt mit Kommandos) mit dem Testsystem zu kommunizieren. Nachdem sie mehrfach nicht verstanden worden waren, gingen diese beiden Probanden jedoch auch zu Kommandos über. Die Erweiterung der Grammatik kann in einem solchen Fall dazu führen, dass die ursprüngliche gewählte Formierungsart der kurzen Phrasen beibehalten werden kann.

Implementierung des „speech-too-early“

Die zeitliche Verzögerung zwischen dem Tonsignal und der Öffnung des Spracherkenners hatte während der explorativen Studie für Verwirrung gesorgt. Sehr häufig wurde versucht zu sprechen, bevor der Spracherkenner geöffnet war. Leider gab es hier nicht viele Änderungsoptionen. Dass der Ton kurz vor der Aktivierung des Spracherkenners abgegeben werden muss, lässt sich nicht ändern, da er sonst mit erkannt werden würde. Allerdings ist es möglich, den Benutzer direkt auf sein Fehlverhalten hinzuweisen. Im Testsystem folgt, wenn der Benutzer zu früh gesprochen hat, grundsätzlich die Systemausgabe „wie bitte?“. Diese ist jedoch in dieser Situation wenig hilfreich, da sie auch ausgegeben wird, wenn der Benutzer undeutlich, zu leise oder zu laut spricht oder wenn er Worte benutzt, die nicht in der Grammatik hinterlegt sind. Die meisten Benutzer folgern aus der Systemausgabe „wie bitte?“, dass sie deutlicher sprechen müssten. Um den Benutzer hier in die richtige Richtung zu lenken, wurde das vom Nuance Spracherkenner angebotene Funktion „speech-too-early“ genutzt. Dieses registriert anhand der nicht vorhandenen Pause zu Beginn einer Äußerung, wenn der Benutzer vor der tatsächlichen Öffnung des Spracherkenners angefangen hat zu sprechen. Im XML-Skript wurde daraufhin festgelegt, dass nach einem Ereignis „speech-too-early“ die Systemausgabe: „Bitte sprechen Sie erst nach dem Signalton“ erfolgen sollte. Diese ist zwar inhaltlich nicht ganz korrekt, da das Problem wie beschrieben darin besteht, dass der Ton eben ein paar Millisekunden zu früh abgegeben wird, die Ausgabe führt aber nach ersten Überlegungen dennoch dazu, dass die Versuchspersonen ein wenig länger warten, bevor sie anfangen

zu sprechen. Außerdem scheint es nicht sinnvoll zu sein, in der Systemausgabe auf das Mikrofon-Symbol statt auf den Ton hinzuweisen. Eine Ausgabe „Bitte sprechen Sie erst, nachdem das Mikrofon eingeblendet wurde“ ist auf keinen Fall ratsam, da sie den Benutzer zwingen würde, auf den Bildschirm zu schauen. Dies sollte bei der Sprachbedienung möglichst selten notwendig sein. Hinzu kommt, dass in der explorativen Studie häufiger auf den Ton als auf das Mikrofon-Symbol geachtet worden war.

Zusätzlicher Ton beim Initial Timeout

Der Ton enthält die Information, ab wann der Spracherkenner geöffnet ist. Der gleiche Inhalt soll durch das Einblenden des Mikrofon-Symbols übermittelt werden. Das Symbol beinhaltet aber noch eine zusätzliche Information durch den Zeitpunkt, an dem es wieder ausgeblendet wird. Dieser entspricht dem Zeitpunkt des Initial Timeout, also genau dem Augenblick, in dem sich der Spracherkenner wieder schließt, wenn der Benutzer nichts gesagt hat (nach vier Sekunden). Dennoch wurde während der explorativen Studie festgestellt, dass mehr Personen auf den Ton als auf das Symbol achten. Um die zusätzliche Information, die das Symbol bietet, auch den Benutzern zugänglich zu machen, die nicht auf das Symbol achten, wurde ein zusätzlicher Ton eingefügt. Dieser erfolgt zum Zeitpunkt des Initial Timeout und wird im Folgenden mit Endton bezeichnet, um ihn von dem bereits bekannten Anfangston zu unterscheiden. Um Verwechslungen vorzubeugen, wurde darauf geachtet, dass der Anfangs- und Endton in ihrer Tonhöhe deutlich verschieden sind. Der Einfachheit halber wurde daher der Ton verwendet, der im Testsystem bei einer Fehlerkennung vor dem „wie bitte?“ ertönt. Er wurde bereits in Kapitel 3.2.2 beschrieben und in der Variante entfernt. Seine Funktion war ohnehin fragwürdig, da der Fehler, auf den er hinweisen sollte, auch durch das folgende „wie bitte?“ hinreichend verdeutlicht wird.

Globales Kommando „Zurück“ hinzugefügt

Analog zur Schaltfläche „Zurück“ (vgl. hierzu Kapitel 4.2.1), wurde auch ein Kommando „Zurück“ eingefügt. Auch bei einem sprachlich geführten Dialog kann also durch „Zurück“ der letzte Schritt rückgängig gemacht werden.

Bei multimodaler Bedienung führt dies zu einer weiteren Problemstellung. Soll bei einem sprachlichen Kommando „Zurück“ der letzte sprachlich ausgeführte Schritt rückgängig gemacht werden oder unabhängig von der Eingabemodalität einfach nur der letzte Schritt? Da davon ausgegangen werden kann, dass der Benutzer immer den letzten von ihm durchgeführten Schritt meint, wurde entschieden, bei „Zurück“ grundsätzlich den

letzten Schritt rückgängig zu machen, unabhängig davon, mit welcher Modalität das „Zurück“ und der letzte Schritt eingegeben wurde.

Dies führt jedoch zu einem Folgeproblem. Sprachlich und graphisch geführter Dialog verlaufen nicht zwingend analog zueinander. Beispielsweise können, seit der Entwicklung der Variante, bei einem sprachlich geführten Dialog die Anfangsbuchstaben bereits in der Listenansicht eingegeben werden, während bei der GUI-Bedienung ein Wechsel zur Filteransicht zwingend notwendig ist. Was passiert nun, wenn der Benutzer in der Listenansicht einen Anfangsbuchstaben mit Sprache eingegeben hat und dann auf die Schaltfläche „Zurück“ klickt? Soll der Buchstabe gelöscht werden? Diese Frage lässt sich eindeutig mit ja beantworten. Soll der Dialog nun aber in der Listenansicht bleiben oder zur Filteransicht springen? Schließlich liegt durch den Klick ein Modalitätswechsel zur GUI-Bedienung vor. Zu diesem Zeitpunkt, also wenn der Klick auf die Schaltfläche „Zurück“ erfolgt, lässt sich aufgrund der momentanen Struktur des Dialogs nicht mehr feststellen, mit welcher Modalität die vorherigen Eingaben gemacht wurden. Daher kann hier nicht je nach Art der Eingabe ein unterschiedlicher Dialogverlauf definiert werden. Folglich muss die Entscheidung, ob zur Filteransicht gewechselt oder in der Listenansicht geblieben wird, grundsätzlich getroffen werden. Es wurde entschieden, die erstgenannte Option umzusetzen: Es wird zur Filteransicht gewechselt. Der Grund dafür war, dass der mit Sprache bedienende Benutzer auch in der Filteransicht buchstabieren, korrigieren und löschen kann, der graphisch bedienende Benutzer in der Listenansicht jedoch nicht.

Löschen des Elements X nach Verneinung der Frage „Meinten Sie X?“

Während der explorativen Studie war festgestellt worden, dass die Benutzer den Grund für Fehlerkennungen häufig nicht verstanden. Insbesondere bei Homophonen oder ähnlichen Phonemen war ihnen nicht klar, dass diese der Grund für die Fehlerkennungen sein könnten. Um nun an dieser Stelle den Benutzern die Eingabe zu erleichtern, wurde folgende Änderung vorgenommen: Verneint ein Benutzer die Frage „Meinten Sie X?“, wird die Stadt X aus der Liste gelöscht. Die Anzahl der Einträge passt sich ebenfalls an.

Abbildung 4.4 zeigt dieses Verhalten beispielhaft am Eintrag „Aalen“. Dieser wurde in der Variante (rechte Abbildung) soeben gelöscht, nachdem die Frage „Meinten Sie Aalen“ verneint worden war. Im Testsystem (linke Abbildung) bleibt der Eintrag jedoch

bestehen, womit es dem Benutzer erschwert wird, seinen gewünschten Zielort („Ahlen“) einzugeben.



Abbildung 4.4: Vergleich Testsystem (links) und Variante (rechts): „Aalen“ noch vorhanden bzw. gelöscht

Neben den oben dargestellten Weiterentwicklungen wurden auch die kleineren Fehler, die in der explorativen Studie aufgetreten waren, behoben. Das Barge-In wurde funktionsfähig gemacht und die Stellen, an denen der Dialog einfach stehen geblieben war, geändert.

Tabelle 4.6 zeigt die Änderungen am Sprachdialog noch einmal in der Übersicht.

Tabelle 4.6: Übersicht Sprachdialog-Änderungen

Testsystem	Variante
Buchstabieren, Korrigieren und Löschen nur in der Filteransicht	Buchstabieren, Korrigieren und Löschen in Filter- und Listenansicht
Prompts	neue und geänderte Prompts
Grammatik beschränkt sich auf wenige Worte und Phrasen	Erweiterung der Grammatik
Prompt „wie bitte?“ bei leerem Erkennungsergebnis, keine Auswertung von „speech-too-early“	Prompt „Bitte sprechen Sie erst nach dem Signalton“ bei speech-too-early
—	Ton bei Timeout

Tabelle 4.6: Übersicht Sprachdialog-Änderungen [Fortsetzung]

Testsystem	Variante
Ton bei Fehlerkennung	—
—	Kommando „Zurück“
—	Bei Verneinung der Frage „Meinten Sie XY?“ wird das entsprechende Element aus der Liste gelöscht.

Während der Entwicklung der beiden Varianten (eine für die GUI, eine für den Sprachdialog) fiel auf, dass sich die beiden auch gut kombinieren ließen. Es gibt keine Änderung, die in zwei verschiedenen Variationen in den Varianten implementiert worden ist. Auch widersprechen sich die Änderungen in Sprachdialog und GUI nicht, so dass sie auch gemeinsam als Variante fungieren können. Daher wurde beschlossen, statt zwei nur eine Variante zu entwickeln, die alle Änderungen umfasst. Diese Entscheidung wurde auch vor dem Hintergrund getroffen, dass im anschließenden Benutzertest bei einer größeren Anzahl an Varianten auch mehr Probanden nötig sein würden.

5 Benutzertest

Die entwickelte, in Kapitel 4.2 vorgestellte Variante soll nun getestet werden. Dabei dient das Testsystem aus Regensburg als Vergleichsobjekt. Im Folgenden werden zunächst die formulierten und zu testenden Hypothesen vorgestellt. Anschließend werden die Methoden zur Überprüfung der Hypothesen, d.h. insbesondere der Versuchsaufbau, die Aufgabenstellung und der Fragebogen, illustriert. Schließlich wird auf die Ergebnisse eingegangen.

5.1 Fragestellung und Hypothesen

Der Benutzertest soll zeigen, ob die vorgenommenen Änderungen am Testsystem tatsächlich Verbesserungen darstellen. Die wichtigsten Punkte dabei sind Effektivität, Effizienz und Akzeptanz der entwickelten Variante. Zu diesen und einigen weiteren Aspekten wurden Hypothesen formuliert, die anhand des Tests überprüft werden sollen.

1. Bei der Variante ist die Task-Completion-Rate höher als beim Testsystem.

Aspekt: Effektivität

Begründung für diese Hypothese: Es wird davon ausgegangen, dass die Benutzer bei der Bedienung der Variante durch das erweiterte Vokabular, die geänderten Prompts und die bessere Selbsterklärungsfähigkeit des Systems häufiger ihr Dialogziel erreichen als beim Testsystem. Dies bedeutet insbesondere, dass es seltener zu einem systemseitigen Abbruch kommt.

2. Bei der Variante werden weniger Dialogschritte benötigt als beim Testsystem.

Aspekt: Effizienz

Begründung für diese Hypothese: Insbesondere durch die Möglichkeit, bei Sprachbedienung auch in der Listenansicht buchstabieren zu können, benötigt der Benutzer vermutlich weniger Dialogschritte als beim Testsystem. Hinzu kommt, dass durch die neuen Beschriftungen der Tasten und Schaltflächen seltener „Fehl-Klicks“ auftreten sollten.

3. Bei der Variante ist die Bedienzeit kürzer als beim Testsystem.

Aspekt: Effizienz

Begründung für diese Hypothese: Auch bei der Bedienzeit sind die benötigten Dialogschritte (vgl. Hypothese 2) von Bedeutung. Nicht immer, aber meistens bedeuten weniger Dialogschritte eine kürzere Bedienzeit. Darüber hinaus spielen aber auch die Anzahl der Timeouts, der systemseitigen Abbrüche und der speech-too-early-Situationen eine Rolle (vgl. Hypothesen 1, 6 und 7). Bei all diesen wird davon ausgegangen, dass sie bei der Variante deutlich seltener auftreten als beim Testsystem. Gegen eine kürzere Bedienzeit bei der Variante spricht die Hypothese 9, die davon ausgeht, dass bei der Variante längere Sätze verwendet werden. Auch die geänderten, zumeist längeren Prompts sollten die Bedienzeit verlängern. Es wird jedoch davon ausgegangen, dass diese beiden Aspekte nur marginale Auswirkungen auf die Gesamtbedienzeit haben werden.

4. Bei der Variante treten weniger Fahrfehler auf als beim Testsystem.

Aspekt: Verkehrssicherheit

Begründung für diese Hypothese: Es wird angenommen, dass bei der Variante der gesamte Sprachdialog „flüssiger“ läuft und folglich weniger Konzentration erfordert als beim Testsystem. Daher kann der Fahraufgabe mehr Aufmerksamkeit gewidmet werden, so dass insgesamt weniger Fahrfehler auftreten.

5. Bei der Variante ist die Akzeptanz des Gesamtsystems höher als beim Testsystem.

Aspekt: Akzeptanz

Begründung für diese Hypothese: Nach [Kan02, S. 22ff.] ist die Akzeptanz eines sprachbasierten Assistenzsystems bedingt durch das Vertrauen, das dem System entgegengebracht wird. Dieses Vertrauen wiederum beruht auf subjektiven Bewertungen hinsichtlich vieler Faktoren, z.B. Nützlichkeit, Zuverlässigkeit und Verstehbarkeit des Systems. Es wird davon ausgegangen, dass aufgrund der vorgenommenen Änderungen das Vertrauen und somit die Akzeptanz der Variante größer ist als die des Testsystems.

6. Bei der Variante treten weniger speech-too-early-Situationen auf als beim Testsystem.

Aspekt: Selbsterklärungsfähigkeit, Verständlichkeit

Begründung für diese Hypothese: Der Sprachdialog macht bei der Variante durch die Systemausgabe „Bitte sprechen Sie erst nach dem Signalton“ darauf aufmerksam, dass zu früh gesprochen wurde. Da die Benutzer daraufhin vermutlich ihr Verhalten ändern, sollte diese Situation insgesamt seltener auftreten.

7. Bei der Variante kommt es seltener zu Initial Timeouts als beim Testsystem.

Aspekt: Selbsterklärungsfähigkeit, Verständlichkeit

Begründung für diese Hypothese: Durch die geänderten Prompts, die nun direkte Handlungshinweise enthalten, wird den Benutzern einiges „Grübeln“ erspart. Dadurch sollte es seltener zu Initial Timeouts kommen. Zusätzlich wird den Benutzern durch den Ton beim Timeout deutlicher, dass diese Situation eingetreten ist, so dass sie beim nächsten Mal ihr Verhalten ändern und rechtzeitig sprechen können.

8. Bei der Variante wird seltener als beim Testsystem etwas gesagt, ohne den Spracherkenner aktiviert zu haben.

Aspekt: Selbsterklärungsfähigkeit, Verständlichkeit

Begründung für diese Hypothese: Wie bei Hypothese 7 erwähnt, erhalten die Benutzer durch den Ton beim Initial Timeout auch einen akustischen Hinweis darauf, dass

die Spracherkennung nicht mehr aktiviert ist. Somit sind die Benutzer bei der Variante besser über den Status des Spracherkenners informiert als beim Testsystem, wo es durch das Ausblenden des Mikrofon-Symbols nur einen visuellen Hinweis darauf gibt. Wenn die Benutzer wissen, dass der Spracherkennung nicht mehr aktiviert ist, werden sie seltener vergessen, ihn durch Drücken der entsprechenden Taste wieder zu aktivieren. Darüber hinaus ist auf dieser Taste bei der Variante anstatt der wenig aussagekräftigen Beschriftung „PTT“ ein selbsterklärendes Mikrofon abgebildet.

9. Bei der Variante werden längere Sätze verwendet als beim Testsystem.

Aspekt: Restriktionen an die Benutzereingabe

Begründung für diese Hypothese: Da durch die Erweiterung der Grammatik mehr Formulierungen abgefangen werden, kommt es seltener vor, dass die Probanden Phrasen sagen, die nicht in der Grammatik hinterlegt sind. Dass die Benutzer wie in der explorativen Studie aus Frustration über Fehlerkennungen von längeren Phrasen zu Kommandos wechseln, sollte daher nicht mehr bzw. seltener vorkommen.

Weitere Fragestellungen

Zusätzlich soll, ohne den Vergleich von Testsystem und Variante, überprüft werden, welche Modalität von den Benutzern im Allgemeinen bevorzugt wird und wie häufig Modalitätswechsel auftreten. Eine wichtige Fragestellung dabei ist, ob sich die Modalitätswahl bei Erhöhung der Schwierigkeit der Aufgabe (Zuschalten der Fahraufgabe, Erhöhung der Geschwindigkeit) ändert.

5.2 Durchführung des Benutzertests

Die oben aufgeführten Hypothesen sollten nun überprüft werden. Dazu wurden unterschiedliche Methoden verwendet. Viele Punkte sollten durch das Verhalten der Versuchspersonen während des Tests und einige durch den anschließenden Fragebogen geklärt werden. Für jede Hypothese wurde die Methode ausgewählt, die für die Gewinnung eindeutiger Ergebnisse am geeignetsten schien.

Insgesamt nahmen 40 Versuchspersonen an dem Benutzertest teil (11 männlich, 29 weiblich). Zwingende Voraussetzung war, dass sie den Führerschein Klasse B bzw. 3

besitzen und Deutsch ihre Muttersprache ist. Ersteres war für die zu bewältigende Fahraufgabe und letzteres für die Sprachbedienung notwendig. Zudem wurde die Einschränkung vorgenommen, dass alle Versuchspersonen Studenten und unter 30 Jahre alt sein müssen. Auf diese Art und Weise sollte sichergestellt werden, dass die Versuchspersonen nicht zu starke Unterschiede in ihrem Verhalten aufweisen, die auf Alter bzw. Ausbildung zurückzuführen sein könnten. Nur so konnte bei der Anzahl von nur 40 Versuchspersonen garantiert werden, dass die Untersuchung verwertbare und signifikant verschiedene Ergebnisse ergibt.

An alle Versuchspersonen wurde ein Vorabfragebogen geschickt, der vor der Durchführung des Tests auszufüllen war (vgl. Anhang D.4). Mit ihm sollten nicht nur vorab bereits alle personenspezifischen Daten wie Geburtsdatum und Studiengang erfasst werden, sondern er diente gleichzeitig als Kontrollinstrument, ob bei den Versuchspersonen Farbenblindheit oder Hörschäden auftraten und sie deshalb von dem Versuch auszuschließen seien.



Abbildung 5.1: Versuchsaufbau Benutzertest

Der Versuchsaufbau ist in Abbildung 5.1 zu sehen. Auf dem frontal vor der Versuchsperson aufgebauten Laptop (in der Abbildung links) läuft die Fahraufgabe - das so genannte Tracking-Tool¹. Es zeigt einen Kreis, der sich auf einer Straße befindet, die sich langsam vorwärts bewegt. Solange sich der Kreis inmitten der Straße befindet, ist er

¹ *Tracking-Tool*. Von der Firma Dubrowsky & Partner, Berlin, entwickelte Software, die versucht, auf möglichst simple Art und Weise die wichtigsten sensomotorischen und kognitiven Anforderungen einer Autofahrt zu simulieren.

grün, kommt er an den Rand (zwischen 0 und 50% des Kreises über dem Fahrbahnrand), wird er gelb. Sobald der Kreis ganz von der Straße abkommt (mehr als 50% des Kreises über dem Fahrbahnrand), nimmt er die Farbe rot an.

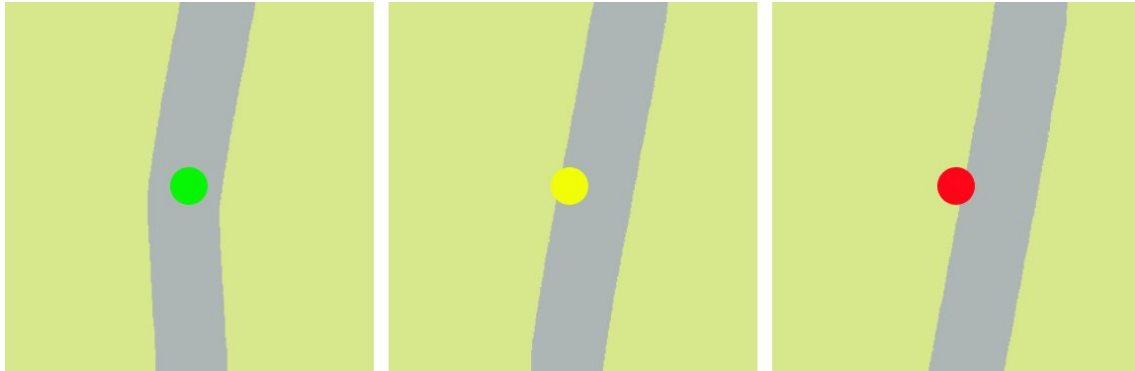


Abbildung 5.2: Tracking-Tool

Diese drei Möglichkeiten sind auch in Abbildung 5.2 zu sehen. Die Steuerung des Tracking-Tools, d.h. die Bewegung des Kreises nach links oder rechts, funktioniert über ein Lenkrad der Firma Microsoft. Gas und Bremse sind nicht zu betätigen, da die Geschwindigkeit voreingestellt ist.

Auf dem zweiten Laptop (in Abbildung 5.1 rechts zu sehen) läuft der Zieleingabe-Dialog. Für die Bedienung stehen den Benutzern eine Maus für die graphische Eingabe sowie ein Headset für die sprachliche Ein- und Ausgabe zur Verfügung. Neben dem Dialogfenster ist das deutsche Buchstabieralphabet eingeblendet, um den Benutzern das Buchstabieren mit Eigennamen zu erleichtern.

Die Aufgabenstellung war in drei Blöcke unterteilt. Zunächst sollten die Versuchspersonen den Zieleingabe-Dialog kennenlernen, ohne parallel die Fahraufgabe bewältigen zu müssen. Im zweiten Block wurde dann die Fahraufgabe in voreingestellter langsamer Geschwindigkeit zugeschaltet. Im dritten Block wurde daraufhin die Geschwindigkeit annähernd verdoppelt.

1. Ohne Fahraufgabe
 - a) GUI
 - b) Sprache
 - c) Wahl

2. Mit langsamer Fahraufgabe
 - a) GUI
 - b) Sprache
 - c) Wahl
3. Mit schneller Fahraufgabe
 - a) Wahl

Block 1 und 2 bestanden aus jeweils 15 Aufgaben, von denen die ersten fünf über die GUI, die zweiten fünf mittels Sprache und die letzten fünf frei nach Belieben bearbeitet werden sollten. Block 3 bestand lediglich aus fünf Aufgaben, die nach Wahl bedient werden konnten. Die Aufgaben ähnelten denen aus der explorativen Studie. Dabei wurden absichtlich Zielorte ausgewählt, die bei Sprachbedienung oder Buchstabieren zu Problemen führen könnten (z.B. Göttingen/Göppingen, Mühlheim/Mülheim, Aalen/Ahlen). Die vollständige Aufgabenstellung des Benutzertests ist dem Anhang D.2 zu entnehmen.

Von den 40 Versuchspersonen testeten 20 die Variante (Experimentalgruppe) und 20 das Testsystem (Kontrollgruppe). Dabei wurde die oben aufgeschlüsselte Reihenfolge a-b-c (GUI-Sprache-Wahl) bei der Hälfte der Versuchspersonen in b-a-c (Sprache-GUI-Wahl) geändert. Damit sollte ausgeschlossen werden, dass Benutzerentscheidungen wie z.B. die Modalitätswahl im Teil c durch Reihenfolgeeffekte² beeinflusst werden.

Ein Versuch dauerte im Durchschnitt 1 Stunde und 15 Minuten. Nach der Begrüßung hatten die Versuchspersonen etwa 3-5 Minuten Zeit zum Kennenlernen des Tracking-Tool. Daraufhin bekamen Sie eine kurze Einführung in die Benutzung des Dialogsystems (vgl. Anhang D.1). Es wurde jedoch darauf geachtet, das System selbst nicht vorzuführen. Sonst hätten im Anschluss einzelne Aspekte wie die Selbsterklärungsfähigkeit des Systems nicht mehr getestet werden können. Für den anschließenden eigentlichen Testteil benötigten die Versuchspersonen zwischen 15 und 45 Minuten. Dabei wurden sie mit einer digitalen Videokamera gefilmt. Anschließend war noch der Fragebogen auszufüllen (vgl. Anhang D.3).

² *Reihenfolgeeffekt*. Die Reihenfolge, in der etwas erlernt oder kennengelernt wird, kann einen entscheidenden Einfluss auf die Akzeptanz und Wiederverwendung des Erlernten haben. Das „Vertraute“, zumeist das zuerst Kennengelernte, wird häufig bevorzugt.

5.3 Ergebnisse

In diesem Kapitel werden die wichtigsten Ergebnisse, die aus dem Benutzertest gewonnen wurden, präsentiert. Zunächst wird auf die formulierten Hypothesen, anschließend auf allgemeinere Fragestellungen ohne den Vergleich Testsystem-Variante eingegangen.

5.3.1 Überprüfung der Hypothesen

Die in Kapitel 5.1 vorgestellten Hypothesen sollen nun überprüft werden. Die Daten, die hierzu herangezogen werden, wurden aus den Vorabfragebögen, den Logfiles, den Videos und den abschließenden Fragebögen gewonnen. Die Auswertung erfolgte mit SPSS®, Version 13, und Microsoft Excel®, Version 2003.

Die erste Hypothese lautete „Bei der Variante ist die Task-Completion-Rate höher als beim Testsystem“. Da bei dem durchgeführten Test ausnahmslos alle Versuchspersonen vor dem Ablauf der vorgegebenen vier Minuten die Aufgaben bewältigen konnten, ließ sich die Task-Completion-Rate nicht mit dem normalerweise üblichen Verfahren (Anzahl der bewältigten Aufgaben/Gesamtanzahl der Aufgaben) berechnen. Als Hilfsmittel wurde daher die Anzahl der benutzer- und systemseitigen Abbrüche hinzugezogen. Das bedeutet, in der hier definierten Task-Completion-Rate zählt eine Aufgabe nur dann als erfolgreich bewältigt, wenn weder der Benutzer noch das System während der Interaktion den Dialog abgebrochen haben. So ergeben sich die Daten in Abbildung 5.3.

Tatsächlich ist die Task-Completion-Rate bei der Variante höher. Dieses Ergebnis ist nach statistischer Berechnung bei einer Irrtumswahrscheinlichkeit³ von 0,009 sehr signifikant⁴. Somit ist davon auszugehen, dass der festgestellte Unterschied nicht nur für die

³ *Irrtumswahrscheinlichkeit.* Die Wahrscheinlichkeit, dass das empirisch gefundene Stichprobenergebnis zustandekommt, obwohl es nicht für die Gesamtpopulation gilt (vgl. hierzu [BD02, S. 29f.]). Hier und bei allen folgenden Beispielen wurde sie mit der SPSS-Funktion „T-Test bei unabhängigen Stichproben“ berechnet.

⁴ Hier findet der nach [BZ05, S. 113] übliche Sprachgebrauch Anwendung: „Aussagen, die mit einer Irrtumswahrscheinlichkeit $p \leq 0,05$ behaftet sind, nennt man signifikant, solche mit einer Irrtumswahrscheinlichkeit $p \leq 0,01$ heißen sehr signifikant und solche mit einer Irrtumswahrscheinlichkeit $p \leq 0,001$ höchst signifikant“.

Stichprobe, sondern auch für die Gesamtpopulation gilt. Dabei ist hier und auch bei den folgenden Fällen zu bedenken, dass die Gesamtpopulation nicht mit der Gesamtbevölkerung Deutschlands gleichzusetzen ist. Die Einschränkungen, die für die Stichprobe vorgenommen wurden, müssen auch für die Gesamtpopulation gelten. Es kann also nur auf das Verhalten von Studenten unter 30 Jahren, die einen Führerschein besitzen, geschlossen werden.

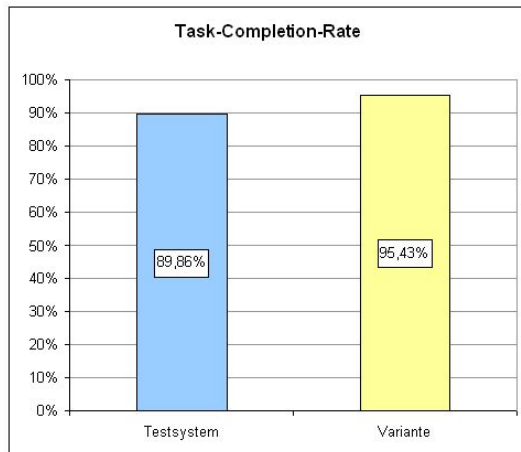


Abbildung 5.3: Task-Completion-Rate

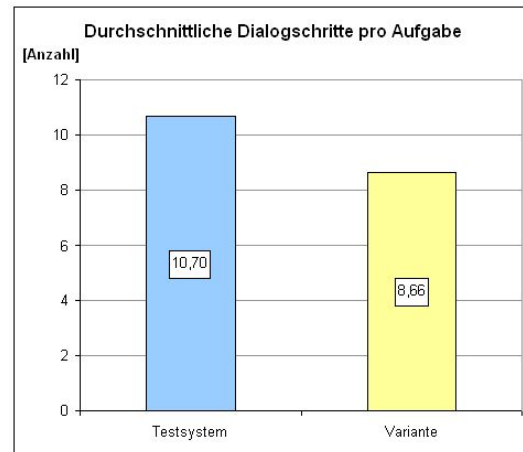


Abbildung 5.4: Dialogschritte

Die zweite Hypothese, „Bei der Variante werden weniger Dialogschritte benötigt als beim Testsystem“, ist ebenfalls anzunehmen. Die Irrtumswahrscheinlichkeit beträgt hier 0,034. Abbildung 5.4 zeigt deutlich, dass die Probanden, die die Variante testeten, um fast 20% weniger Dialogschritte benötigten als ihre Kommilitonen, die das Testsystem benutzen sollten.

Die dritte Hypothese lautete: „Bei der Variante ist die Bedienzeit kürzer als beim Testsystem“. In Abbildung 5.5 ist deutlich zu sehen, dass die Versuchspersonen, die die Variante testeten, tatsächlich weniger Zeit für die Aufgaben benötigten. Dabei fällt auf, dass der Unterschied beim ersten Versuchsblock („ohne Fahrt“) sichtlich größer ist als beim zweiten („bei langsamer Fahrt“). Beim dritten Versuchsblock ist der Unterschied bereits nicht mehr signifikant (zum Vergleich der Irrtumswahrscheinlichkeiten vgl. Tabelle 5.1).

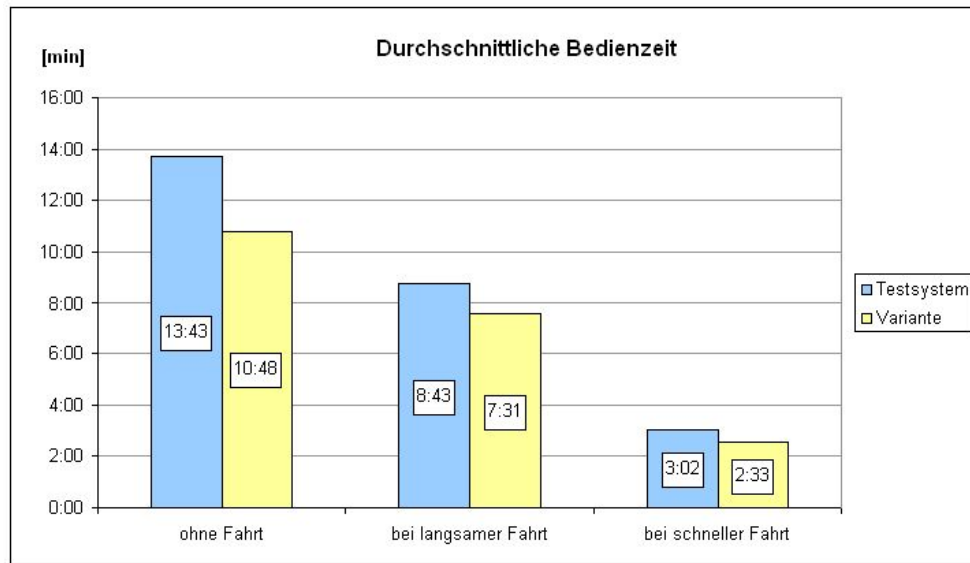


Abbildung 5.5: Durchschnittliche Bedienzeit über alle Aufgaben

Tabelle 5.1: Signifikanz bei Durchschnittlicher Bedienzeit

	Testsystem	Variante	Irrtumswahrscheinlichkeit	Signifikanz
Bedienzeit ohne Fahrt	13:43 min	10:48 min	$p = 0,003$	sehr signifikant
Bedienzeit bei langsamer Fahrt	8:43 min	7:31 min	$p = 0,049$	signifikant
Bedienzeit bei schneller Fahrt	3:02 min	2:33 min	$p = 0,178$	nicht signifikant

Die naheliegendste Erklärung für dieses Phänomen ist, dass die Vorzüge der Variante hauptsächlich in den Bereichen „Selbsterklärungsfähigkeit“ und „Natürlichkeit des Dialogs“ liegen (vgl. hierzu Kapitel 2.3.3). Dadurch wird sie insbesondere für den Anfänger leichter und mit weniger Zeitaufwand benutzbar. Auch die Versuchspersonen waren schließlich zu Beginn des Versuchs, d.h. bei Block 1, „Anfänger“ in dem Sinne, dass sie das System noch nicht kannten. Daher fiel der Unterschied zwischen Variante und Testsystem hier mehr ins Gewicht als bei Block 2 und 3, wo die Versuchspersonen schon geübt waren.

Der „Lerneffekt“ der Versuchspersonen ist daran zu erkennen, dass sie, unabhängig von Variante oder Testsystem, für den zweiten Block insgesamt weniger Zeit benötigten als für den ersten - und das, obwohl der zweite Block durch das Zuschalten der langsamen Fahraufgabe deutlich anspruchsvoller war. Das ist anhand von Abbildung 5.6 noch einmal gut nachvollziehbar. Der prozentuale Anteil des dritten Blocks („schnelle Fahrt“) ist selbstverständlich nur deshalb so gering, weil hier nur fünf (statt 15) Aufgaben zu bewältigen waren.

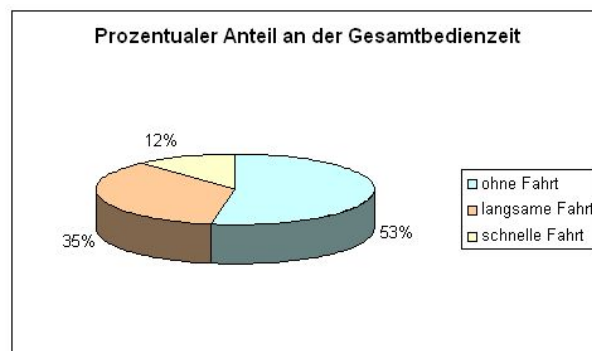


Abbildung 5.6: Prozentualer Anteil der drei Blöcke an der Gesamtbedienzeit

Die Ergebnisse zur vierten Hypothese, „Bei der Variante treten weniger Fahrfehler auf als beim Testsystem“, sind in den Abbildungen 5.7 und 5.8 beispielhaft an den „gelben Fehlern“ dargestellt. Mit „gelbem Fehler“ ist das Ereignis gemeint, wenn sich der Kreis des Tracking-Tools gelb färbt, weil er an den Rand der Fahrbahn kommt. Die geringere Anzahl der Fehler bei der Variante lässt zwar einen signifikanten Unterschied vermuten, hält einer genaueren Prüfung jedoch nicht stand. Betrachtet man die Dauer der Fehler, d.h. die Zeit, während der der Kreis des Tracking-Tools gelb blieb, beträgt der Unterschied nur noch wenige Millisekunden. Auch die Irrtumswahrscheinlichkeiten 0,482 bei der Anzahl der Fehler bzw. 0,808 bei der Dauer bestätigen, dass diese Hypothese zu verwerfen ist. Das gleiche gilt für die „roten Fehler“⁵. Insgesamt hätte bei der vorliegenden Soft- und Hardware ohnehin keine Aussage zur Verkehrssicherheit gemacht werden können. Dafür ist die mit dem Tracking-Tool nachgestellte Fahrsituation zu weit von der Realität, einer tatsächlichen Autofahrt, entfernt. Zudem dürfte auch die Bedienung mit der Maus deutlich verschieden sein von der Interaktion mit einem Touchscreen, wie er im Auto zu finden wäre.

⁵ Auf eine graphische Darstellung der „roten Fehler“ in Form eines Diagramms wurde wegen des geringen Informationsgehaltes verzichtet.

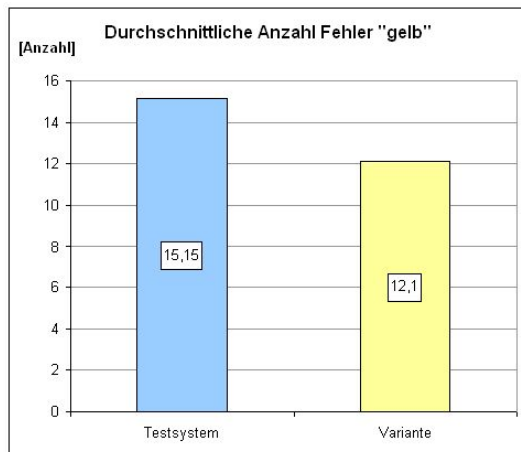


Abbildung 5.7: Anzahl der „gelben“ Fehler

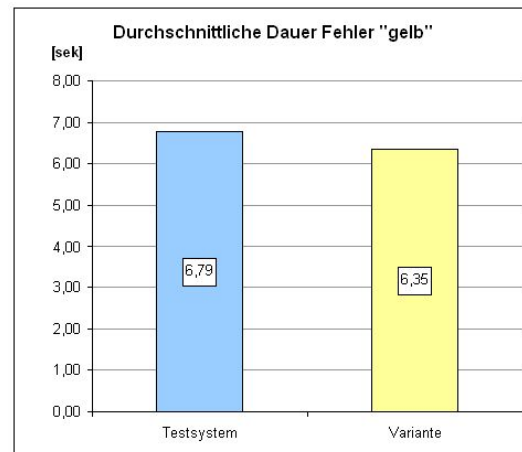


Abbildung 5.8: Dauer der „gelben“ Fehler

Die fünfte Hypothese lautete „Bei der Variante ist die Akzeptanz des Gesamtsystems höher als beim Testsystem“. Ein Großteil des Fragebogens, der in Anhang D.3 zu finden ist, zielte darauf ab, eben diesen Aspekt zu messen. Der letzte Fragenblock mit der Überschrift „Wie würden Sie das System aus Ihrer persönlichen Sicht bewerten?“ wurde zu diesem Zweck aus der Diplomarbeit von Jana Meyer übernommen, die sich eigens mit der Entwicklung eines Fragebogens zur Messung der Akzeptanz auseinandergesetzt hatte (vgl. [Mey02]). Dabei sollten die Versuchspersonen Aussagen wie „Dieses System ist eine wertvolle Unterstützung.“ auf einer fünf-stufigen Skala von „Stimmt gar nicht“ bis „Stimmt völlig“ beurteilen. Den einzelnen Punkten wurden Zahlen von -2 (Stimmt gar nicht) bis +2 (Stimmt völlig) zugeordnet⁶. War die einzuschätzende Aussage negativ formuliert, wurden die Zahlen selbstverständlich umkodiert. Aus dem Mittelwert der einzelnen Antworten ergab sich eine Variable: die Akzeptanz. Sie wurde für das Testsystem und die Variante ermittelt und ist in Abbildung 5.9 dargestellt. Dabei kann die Variable Akzeptanz hier einen Wert von -2 bis +2 annehmen.

Das bedeutet zunächst, dass beide Systeme grundsätzlich eher positiv bewertet wurden, da ihr jeweiliger Akzeptanz-Wert größer als 0 ist. Und tatsächlich hat die Variante, wie in der oben formulierten Hypothese angenommen, einen größeren Wert als das Test-

⁶ Hier wird davon ausgegangen, dass intervallskalierte Daten vorliegen, d.h. dass die Abstände zwischen den einzelnen Punkten gleich groß sind. Nur so kann mit den Daten eine Mittelwertberechnung durchgeführt werden. Auch wenn diese Vorgehensweise in der Fachliteratur umstritten ist, stellt sie die „übliche Forschungspraxis“ dar. [BD02, S. 74]

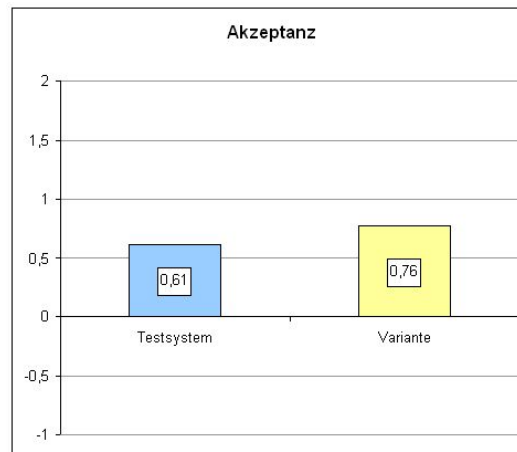


Abbildung 5.9: Akzeptanz

system. Allerdings ist dieses Ergebnis bei einer Irrtumswahrscheinlichkeit von 0,432 nicht signifikant.

Der Fragebogen liefert zu diesem Aspekt noch einige zusätzliche Informationen. Es wurde die offene Frage gestellt, was den Versuchspersonen an der GUI- bzw. Sprachbedienung am besten gefallen und was sie am meisten gestört hat (vgl. hierzu Anhang D.3). In Bezug auf die GUI-Bedienung unterschieden sich die Antworten der Benutzer des Testsystems nicht von denen der Benutzer der Variante. Für gut geheißen wurde die schnelle und direkte Bedienung über die GUI. Schlecht fanden die meisten Versuchspersonen die Anordnung der Pfeile als Navigationstasten. Sie wurde als „unübersichtlich“ und „umständlich“ empfunden. Vermutlich hätte hier ein Scrollbalken, wie er z.B. bei Microsoft Word[®] oder Excel[®] üblich ist, deutlich mehr Zuspruch bei den Benutzern gefunden.

Bei der Sprachbedienung waren sich die Versuchspersonen nur hinsichtlich der positiven Eigenschaften einig. Es wurde für gut befunden, dass der Zielort direkt benannt werden kann und dass die Ablenkung von der Fahraufgabe nur gering ist. Bei den negativen Aspekten der Sprachbedienung zeigte sich hingegen ein deutlicher Unterschied zwischen den Benutzern, die das Testsystem verwendeten, und jenen, die die Variante testeten (vgl. hierzu Abbildung 5.10).

Die Benutzer des Testsystems beanstandeten am häufigsten die Fehlerkennungen (12 Vpn⁷) und deutlich seltener die langen Systemantworten (5 Vpn). Bei den Benutzern der

⁷ Vpn. Abkürzung für Versuchspersonen.

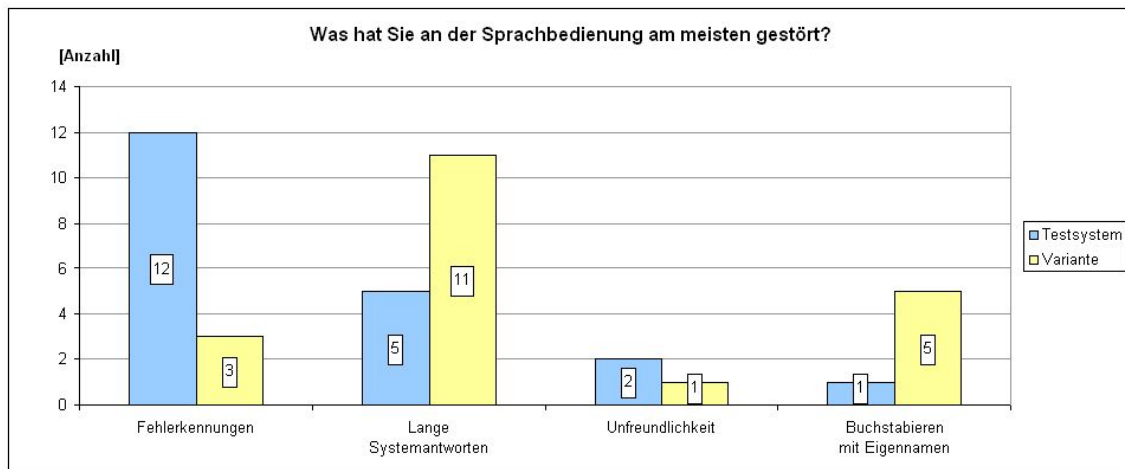


Abbildung 5.10: Negative Aspekte bei der Sprachbedienung

Variante hingegen ist die Antwortverteilung gegenläufig - 11 Versuchspersonen kritisieren die langen Systemantworten und nur 3 die Fehlerkennungen. Dieser Unterschied ist mit einer Irrtumswahrscheinlichkeit von 0,014⁸ signifikant. Hier zeigt sich deutlich die Auswirkung der vorgenommenen Änderungen an der Variante. Durch die Erweiterung der Grammatik konnte die Anzahl der Fehlerkennungen gesenkt werden⁹. Die Änderung der Prompts jedoch hatte scheinbar mehrere erkennbare Auswirkungen. Zunächst einmal zeigte sich der als positiv zu bewertende Aspekt, dass der Dialogablauf besser verstanden wurde und somit die benötigte Bedienzeit verkürzt werden konnte. Zusätzlich führte die Änderung (und zumeist Verlängerung) der Prompts jedoch dazu, dass sich mehr Versuchspersonen über deren „unnötige“ Länge beschwerten. Als besonders wichtig wird in diesem Zusammenhang angesehen, dass die meisten Versuchspersonen bei ihrer Antwort hinzufügten: „wenn ich schon weiß, wie’s geht“ oder „selbst, als ich schon genug geübt hatte“. Das deutet darauf hin, dass die Versuchspersonen die langen Systemausgaben zunächst, d.h. zu Beginn des Versuchs, nicht als nervig empfanden. Erst als sie mit dem System vertraut waren, beurteilten sie die langen Prompts als „lästig“ und „zu lang“. Hier zeigt sich ein Aspekt, der bereits in Kapitel 2.3.3 erwähnt wurde. Dort heißt es, ein multimodales Dialogsystem sollte „Sufficiency of adaptation to user

⁸ Da es sich hier um nominalskalierte Daten handelt, erfolgte die Berechnung mit dem Chi-Quadrat-Test nach Pearson.

⁹ Hierbei sollte bedacht werden, dass die Anzahl der Fehlerkennungen u.a. von der Größe der Liste abhängig ist. Bei der hier implementierten Liste mit 1328 Einträgen sind Fehlerkennungen sehr selten. Steigt die Anzahl der Listeneinträge, wird sich auch die Zahl der Fehlerkennungen erhöhen.

differences“ vorweisen können. Diese Anpassungsfähigkeit an Benutzer mit unterschiedlichen Voraussetzungen/Erfahrungen wird also auch in dem vorliegenden Test in Form einer Anfänger- und Expertenversion gefordert.

Die sechste Hypothese, „Bei der Variante treten weniger speech-too-early-Situationen auf als beim Testsystem“, kann eindeutig angenommen werden. Die dazugehörigen Daten sind in Abbildung 5.11 zu sehen. Das Ereignis „speech-too-early“ trat bei der Variante insgesamt 77 Mal weniger auf als beim Testsystem, pro Versuchsperson (es waren 20 in jeder Gruppe) also fast vier Mal weniger. Die Irrtumswahrscheinlichkeit liegt bei 0,029, womit das Ergebnis als signifikant bezeichnet werden kann.

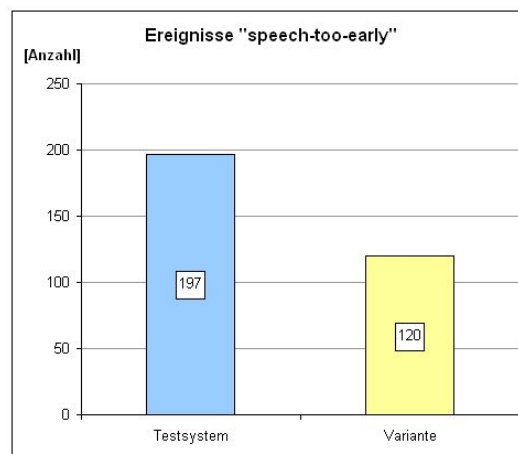


Abbildung 5.11: Häufigkeit des Ereignisses „speech-too-early“

Ebenso kann die siebte Hypothese, „Bei der Variante kommt es seltener zu Initial Timeouts als beim Testsystem“, angenommen werden. Tatsächlich kommt der Initial Timeout bei der Variante fast nur halb so oft vor wie beim Testsystem (vgl. Abbildung 5.12). Das Ergebnis ist bei einer Irrtumswahrscheinlichkeit von 0,038 signifikant.

Die achte Hypothese lautete „Bei der Variante wird seltener als beim Testsystem etwas gesagt, ohne den Spracherkenner aktiviert zu haben“. Das Ergebnis ist in Abbildung 5.13 dargestellt. Zwar ist hier ein deutlicher Unterschied zu sehen, der bei einer Irrtumswahrscheinlichkeit von 0,012 auch als signifikant zu bezeichnen ist, jedoch sollte bedacht werden, dass diese Situation insgesamt bei 40 Versuchspersonen nur 32 Mal aufgetreten ist. Daher ist der Stellenwert dieser Hypothese, wenn man von einem solchen sprechen kann, gering.



Abbildung 5.12: Häufigkeit eines Initial Timeout

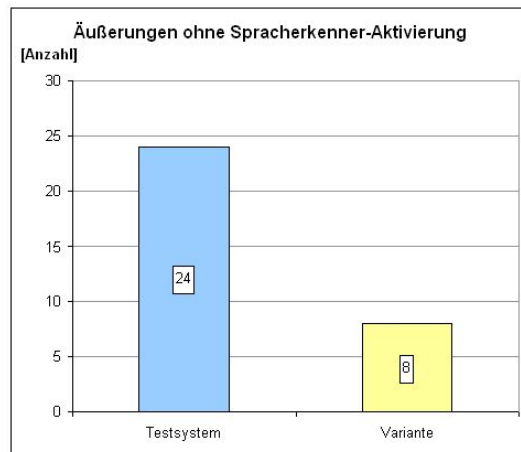


Abbildung 5.13: Häufigkeit einer Äußerung ohne Spracherkenner-Aktivierung

Die neunte Hypothese, „Bei der Variante werden längere Sätze verwendet als beim Testsystem“, wurde verworfen. Hierzu wurden jedoch keine empirisch gewonnenen Daten hinzugezogen, sondern lediglich der Eindruck der Versuchsleiterin während der Durchführung der Tests. Die Auswertung jeder einzelnen Äußerung der Versuchspersonen, d.h. das Zählen der verwendeten Wörter, hätte einen enorm großen zeitlichen Aufwand bedeutet, der nur dann investiert worden wäre, wenn die ersten Eindrücke viel versprechend gewesen wären. Es zeigte sich jedoch, wie anhand des Fragebogens festzustellen ist, dass nur eine Versuchsperson überhaupt versucht hatte, den Dialog mit nahegehend natürlichsprachlichen - und somit längeren - Äußerungen zu führen. Auch diese Person ging jedoch im Laufe des Versuchs zu kurzen Kommandos über, wie sie auch alle anderen Versuchspersonen verwendet hatten. Die Begründungen hierfür variierten von „Geht schneller“ (17 Vpn) über „Vermutung, dass so weniger Fehlerkennungen auftauchen“ (19 Vpn) bis zu „Man spricht ja nur mit einer Maschine“ (4 Vpn).

5.3.2 Weitere Ergebnisse

Im nun folgenden Teil werden Ergebnisse vorgestellt, die unabhängig von den formulierten Hypothesen gefunden wurden. Auf einen Vergleich zwischen Testsystem und Variante wird in diesem Abschnitt verzichtet, d.h. die vorliegenden Daten beziehen sich grundsätzlich auf alle 40 Versuchspersonen. Statt dessen soll das Hauptaugenmerk auf der Modali-

tätswahl / den Modalitätswechseln liegen. Insbesondere interessiert die Frage, ob diese in ähnlicher Art und Anzahl in allen drei Versuchsblöcken vorkamen.

Abbildung 5.14 zeigt, zu welcher Modalität sich die Versuchspersonen zu Beginn der Aufgaben entschieden. Hier werden selbstverständlich nur die 15 Aufgaben (die letzten fünf jedes Aufgabenblocks) betrachtet, bei denen, wie in Kapitel 5.2 beschrieben, die Modalität frei gewählt werden konnte.

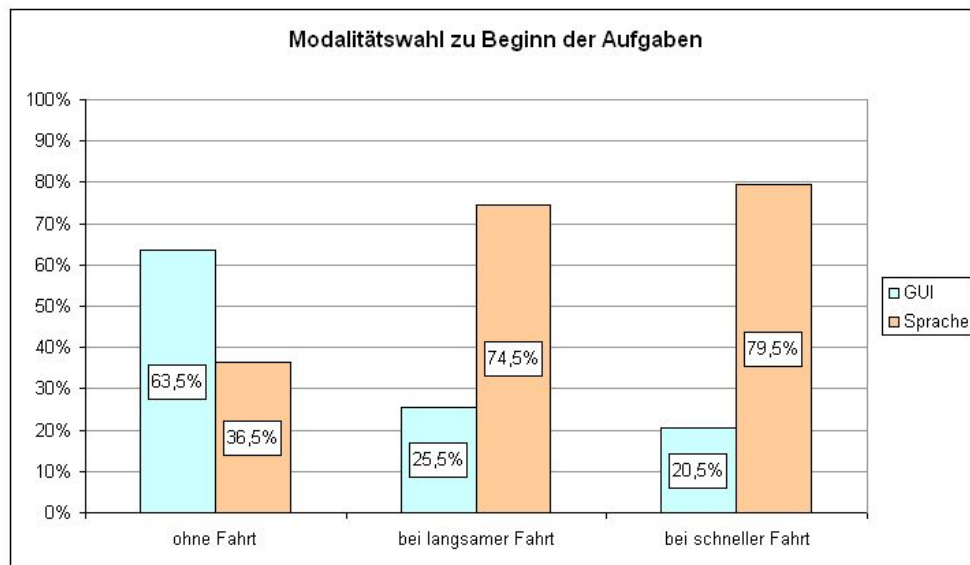


Abbildung 5.14: Modalitätswahl zu Beginn der Aufgaben

Die dargestellten Ergebnisse überraschen nicht. Je wichtiger es wurde, dass die Augen sich auf die Fahraufgabe richten, desto eher entschieden sich die Versuchspersonen für die Modalität Sprache. Am deutlichsten ist der „Sprung“ von Block 1 („ohne Fahrt“) zu Block 2 („bei langsamer Fahrt“) zu sehen. Die Tatsache, dass mit einem Mal zwei Dinge parallel gemeistert werden mussten, scheint dazu geführt zu haben, dass die Sprache als diejenige Modalität, die nur sehr wenig visuelle Aufmerksamkeit benötigt, bevorzugt wurde.

Auch hier wurde mit Hilfe der SPSS-Funktion „T-Test bei abhängigen Stichproben“ ein Signifikanztest durchgeführt. Der Unterschied der Werte von Block 1 und Block 2 ist mit einer Irrtumswahrscheinlichkeit kleiner 10^{-4} höchst signifikant. Das bedeutet also, dass nicht nur in der gewählten Stichprobe, sondern auch in der Gesamtpopulation die Modalitätswahl derart aussieht. Der Unterschied zwischen Block 2 und Block 3 ist nicht signifikant.

Abbildung 5.15 zeigt die vorgenommenen Modalitätswechsel aller Versuchspersonen. Auch sie beziehen sich nur auf die insgesamt 15 Aufgaben, bei denen frei nach Wahl bedient werden konnte.

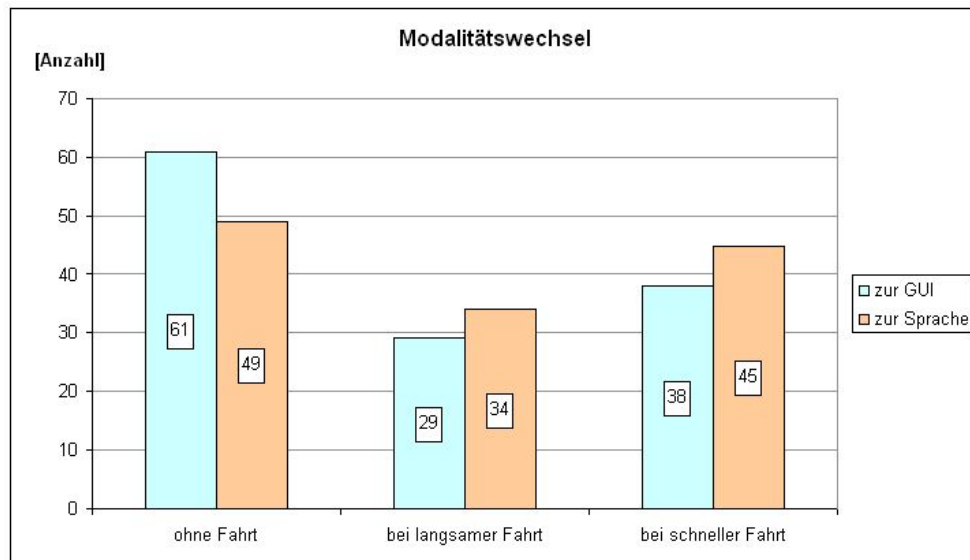


Abbildung 5.15: Anzahl der Modalitätswechsel

Auffällig ist hierbei, dass sich die Daten analog zu jenen aus Abbildung 5.14 verhalten. Während die Fahraufgabe bewältigt werden muss, egal ob langsam oder schnell, wird häufiger zur Sprache und ohne Fahraufgabe häufiger zur GUI gewechselt. Dieses Ergebnis ist mit einer Irrtumswahrscheinlichkeit von 0,018 signifikant.

Zudem ist zu erkennen, dass insgesamt nur wenige Modalitätswechsel auftreten. Bei einer Anzahl von 40 Versuchspersonen gab es während des ersten Blocks („ohne Fahrt“) nur 110 Modalitätswechsel. Das entspricht 2,75 Wechseln pro Person bei fünf zu bewältigenden Aufgaben. Während des zweiten und dritten Versuchsblocks waren es noch weniger. Das lässt darauf schließen, dass die Benutzer generell während eines laufenden Dialogs selten die Modalität wechseln. Tun sie es doch, liegt das zumeist daran, dass sie meinen, mit der gewählten Modalität das Dialogziel nicht bzw. nicht so schnell wie mit der anderen Modalität erreichen zu können. Nach dem Test gaben außerdem ca. 10 Versuchspersonen im persönlichen Gespräch mit der Versuchsleiterin an, dass sie „während der Fahrt“ zur Modalität Sprache gewechselt haben, als ihnen auffiel, dass sie durch die häufigen Blickabwendungen viele Fahrfehler machen.

6 Fazit und Ausblick

Das Dialogsystem zur Zielorteingabe wurde an entscheidenden Stellen verändert. Grundlage für die Änderungen waren die Ergebnisse einer ersten explorativen Studie, die mit acht Versuchspersonen durchgeführt wurde. Diese Ergebnisse beinhalteten neben schwer verständlichen Beschriftungen und verwirrenden Anordnungen auf der graphischen Oberfläche insbesondere Probleme bei der sprachlichen Interaktion mit dem Testsystem (gemeint ist die Regensburger Ausgangsversion des Dialogsystems). Bei der Entwicklung der Variante wurde versucht, für jedes der erkannten Probleme eine Lösung zu finden und somit den gesamten Dialogablauf für den Benutzer zu vereinfachen.

Um festzustellen, ob die verwirklichten Änderungen tatsächlich Verbesserungen darstellen, wurde anschließend ein Benutzertest durchgeführt. Damit sollte auch sichergestellt werden, dass durch die Veränderung des Dialogs keine neuen, vorher nicht vorhandenen Probleme auftreten.

Die Ergebnisse waren überwiegend positiv. Die Task-Completion-Rate konnte erhöht und die Anzahl der benötigten Dialogschritte gesenkt werden. Damit einher ging die deutliche Senkung der Bedienzeit. Dieser Punkt gilt als besonders wichtig. Durch die kürzere Bedienzeit wird die Verkehrssicherheit erhöht, da die Ablenkung von der Fahraufgabe insgesamt abnimmt. Zudem wird ein System eher vom Nutzer akzeptiert, wenn er sein Dialogziel schneller erreichen kann.

Die Anzahl und Dauer der Fahrfehler blieb jedoch konstant. Dies ist nicht weiter verwunderlich, wenn man bedenkt, dass die Nebenaufgabe an sich (Bedienung eines Dialogsystems zur Zielorteingabe) gleich blieb. Die eigentliche Belastung der Versuchspersonen, nämlich zwei Dinge gleichzeitig durchzuführen, blieb erhalten. Daher kam es auch zu vergleichbar vielen Fahrfehlern.

Bei der Bedienung des Testsystems führten das frühe Tonsignal beim Start des Erkenners, der nicht eindeutige Prompt „Wie bitte?“ und der fehlende akustische Hinweis

auf den Initial Timeout zu Bedienproblemen. Sehr häufig wurde zu früh versucht zu sprechen („speech-too-early“) oder ohne dass der Spracherkenner aktiviert worden war. Zudem kam es oft zum Initial Timeout. Die Häufigkeit des Auftretens dieser Probleme konnte bei der Variante deutlich gesenkt werden. Hier waren die Einführung eines Tonsignals beim Initial Timeout und der Prompt „Bitte sprechen Sie erst nach dem Tonsignal“, der im Falle eines „speech-too-early“-Ereignisses abgegeben wurde, die wesentlichen Fortschritte.

Trotz dieser Verbesserungen konnte die Akzeptanz des Systems nicht signifikant erhöht werden. Dabei muss jedoch bedacht werden, dass die Versuchspersonen das Testsystem und die Variante nicht vergleichend beurteilen, sondern lediglich eine Aussage zu dem von ihnen getesteten System machen sollten. Insgesamt waren sich die beiden Systeme vermutlich noch zu ähnlich, als dass sie unterschiedliche Bewertungen bei der Akzeptanz erhalten hätten. Grundsätzlich müssen Verbesserungen am Dialog schließlich nicht zwingend eine Erhöhung der Akzeptanz nach sich ziehen.

Neben den vergleichenden Ergebnissen zu Testsystem und Variante wurden beim Benutzertest noch einige weitere Faktoren beobachtet. Beispielsweise hat sich gezeigt, dass die Modalität „Sprache“ während der Fahrt deutlich bevorzugt wird. Dies ist wünschenswert, weil durch die Nutzung der Modalität „Sprache“ die Ablenkung vom „Fahren“ minimiert wird. Dennoch ist auch die Bedienung über die GUI zur Korrektur falsch erkannter Spracheingaben oder bei Benutzern, die ihr grundsätzlich den Vorzug geben, wichtig. Damit wurde bekräftigt, dass die Entwicklung multimodaler Dialogsysteme für das Kraftfahrzeug der richtige Weg ist.

Die entwickelte Variante wird in ihrer momentanen Form noch nicht in ein Navigations- oder Fahrerinformationssystem der Firma Robert Bosch GmbH integriert werden. Dazu fehlt dem System die notwendige Reife. Allerdings wird das gesamte Wissen, das mit dieser Arbeit entstand, bei der weiteren Entwicklung sprachbasierter und multimodaler Informationssysteme berücksichtigt werden. Insbesondere die Ergebnisse des Benutzertests werden begutachtet und verwertet werden. Dadurch wird diese Arbeit helfen, die Entwicklung innovativer und benutzerfreundlicher Fahrerinformationssysteme einen kleinen Schritt voran zu bringen.

Auch wird die Arbeit dazu anregen, die multimodalen Bedienkonzepte, die der Variante und somit dem Zieleingabedialog zugrunde liegen, zu überdenken und ggf. auf andere Bereiche auszuweiten. Hier tun sich neue interessante Fragestellungen auf: Können und

möchten die Benutzer mehrere Applikationen parallel bedienen, wenn ihnen mehrere Modalitäten zur Verfügung stehen? Gibt es Einzelsysteme, die grundsätzlich nur über eine Modalität bedient werden sollten? Wird die Sprachbedienung von Benutzern unterschiedlichen Alters ähnlich verwendet und akzeptiert? Auf diesem Gebiet kann noch viel geforscht und muss noch viel hinzugelernt werden.

Die Zukunft sprachbasierter und multimodaler Fahrerinformations- und -assistenzsysteme ist noch ungewiss. Zwar wird die Entwicklung derartiger Systeme von Automobilherstellern und -zulieferern stetig vorangetrieben, jedoch lässt sich die Nachfrage nach solchen Produkten nur schwer prognostizieren. Zudem hängt der Erfolg der Systeme nicht nur von gut gestalteten Dialogen, sondern auch von neuer Funktionalität ab. Welche neuartigen Funktionen bietet das System dem Käufer und Autofahrer? Hierauf begründet sich zumeist der Kaufanreiz. Dabei ist der sinnvoll und funktional gestaltete Dialog allerdings eine wichtige Voraussetzung zur Nutzung dieser Systeme. Und letztendlich sind es die Autofahrer, die entscheiden, ob sich die Bediensysteme im Kraftfahrzeug zukünftig mit Sprache, multimodal oder altbewährt über graphisch-visuelle Schnittstellen bedienen lassen.

Anhang

Anhang A

Buchstabieren

A.1 Buchstabieralphabet

Tabelle A.1: Deutsches und NATO-Buchstabieralphabet

Deutsch	NATO	Deutsch	NATO
Anton	Alpha	Paula	Papa
Berta	Bravo	Quelle	Quebec
Caesar	Charlie	Richard	Romeo
Dora	Delta	Siegfried	Sierra
Emil	Echo	Theodor	Tango
Friedrich	Foxtrott	Ulrich	Uniform
Gustav	Golf	Viktor	Victor
Heinrich	Hotel	Wilhelm	Whisky
Ida	India	Xanthippe	X-Ray
Julius	Juliett	Ypsilon	Yankee
Kaufmann	Kilo	Zacharias	Zulu
Ludwig	Lima	Ärger	
Martha	Mike	Ökonom	
Nordpol	November	Übermut	
Otto	Oscar		

A.2 Verwendete Wörter

Tabelle A.2: In der Grammatik der Variante definierte Wörter zum Buchstabieren

Deutsch	NATO	Weitere
Anton	Alpha	Anna
Berta	Bravo	Beta
Caesar	Charlie	Christa
Dora	Delta	Dieter
Emil	Echo	Egon Erich
Friedrich	Foxtrott	
Gustav	Golf	Gerda
Heinrich	Hotel	Hans Helmut
Ida	India	Inge
Julius	Juliett	
Kaufmann	Kilo	Konrad
Ludwig	Lima	Land
Martha	Mike	
Nordpol	November	Norbert
Otto	Oscar	
Paula	Papa	Peter
Quelle	Quebec	
Richard	Romeo	Rudolf
Siegfried	Sierra	Susi
Theodor	Tango	
Ulrich	Uniform	Udo Uta Ulla
Victir	Victor	
Wilhelm	Whisky	
Xanthippe	X-Ray	
Ypsilon	Yankee	
Zacharias	Zulu	Zeppelin
Ärger		
Ökonom		
Übermut		Übel

Anhang B

Dialogabläufe des Testsystems

B.1 Legende

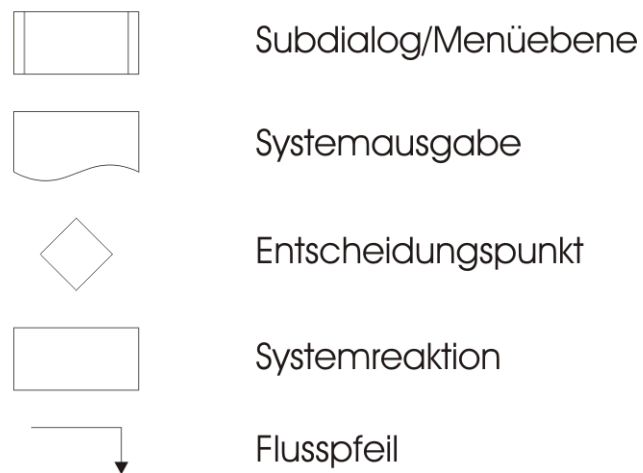


Abbildung B.1: Testsystem: Flowchart Legende

B.2 Hauptmenü

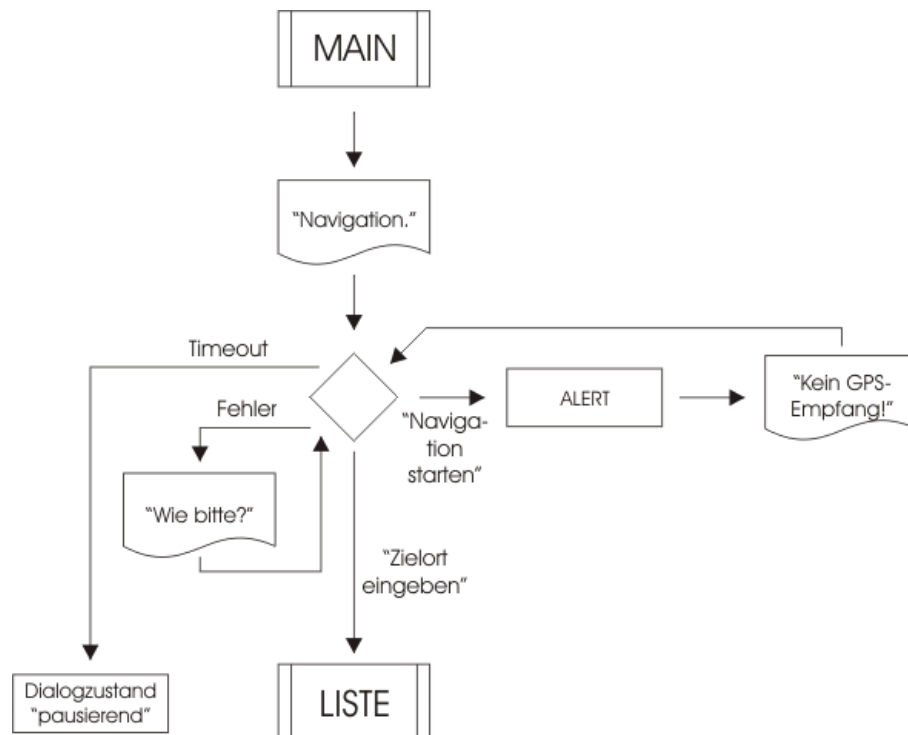


Abbildung B.2: Testsystem: Flowchart Hauptmenü

B.3 Listenansicht

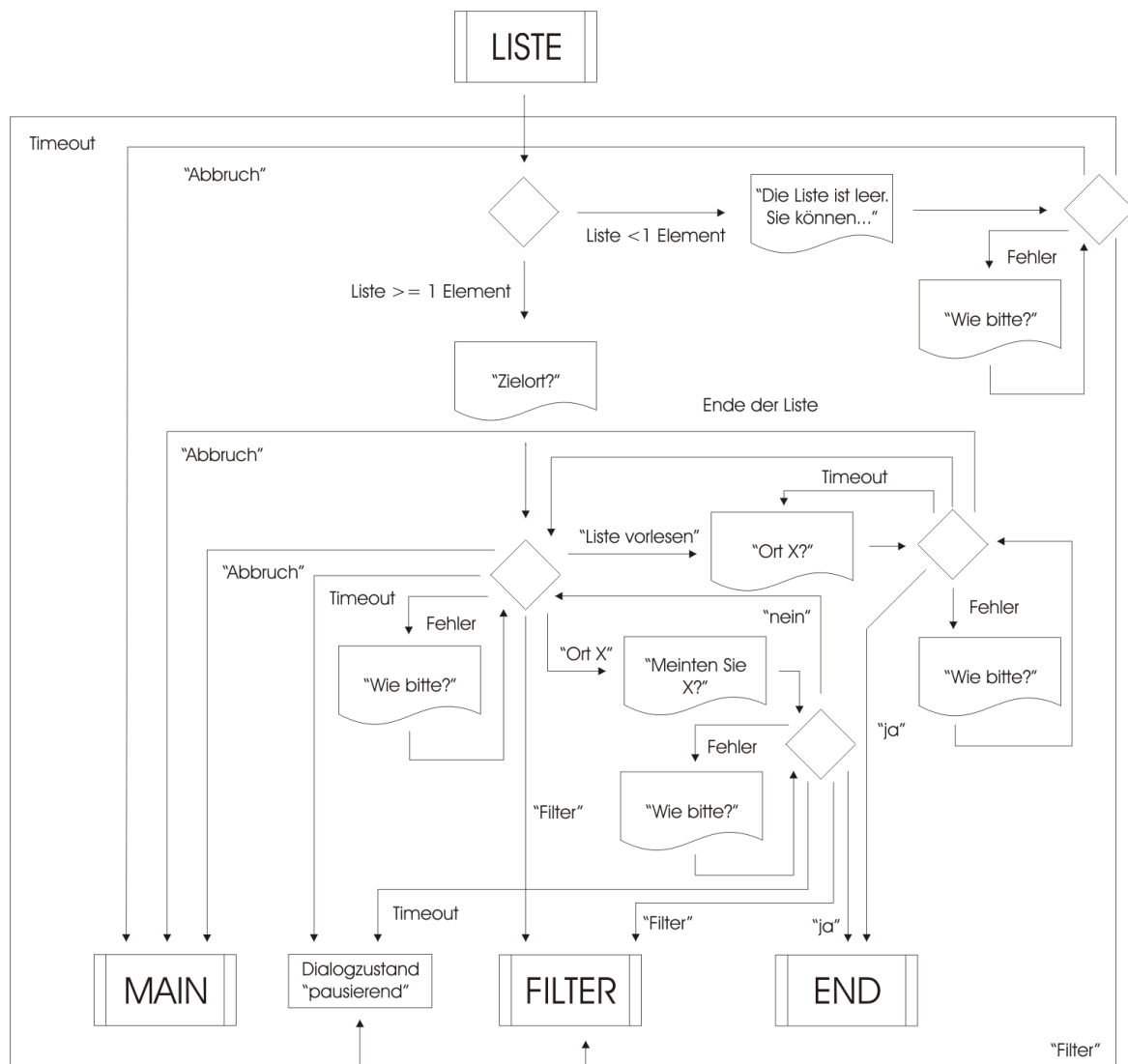


Abbildung B.3: Testsystem: Flowchart Listenansicht

B.4 Filteransicht

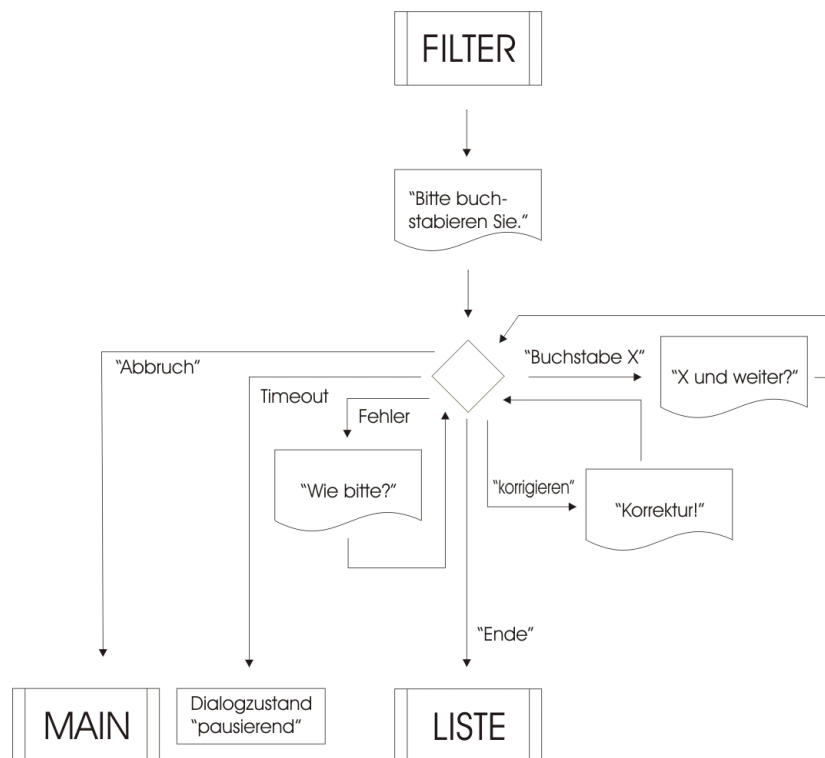


Abbildung B.4: Testsystem: Flowchart Filteransicht

Anhang C

Explorative Studie

Auf den nächsten Seiten sind die Aufgabenstellung und der Fragebogen zur explorativen Studie zu finden.

Sehr geehrter Versuchsteilnehmer,

zunächst wird Ihnen die Versuchsleiterin anhand von zwei Beispielen das Testsystem vorführen. Danach bearbeiten Sie bitte die unten stehenden Aufgaben in der vorgegebenen Reihenfolge. Dafür haben Sie so viel Zeit, wie Sie wollen bzw. benötigen. Mehrfachversuche sind durchaus möglich und auch nicht ungewöhnlich. Schließlich werden nicht Sie getestet, sondern das System.

Viel Spaß mit den Aufgaben und vorweg schon mal vielen Dank!

Übungstrials

Bedienung mit der Maus: **Berlin**

Bedienung mittels Sprache: **Dresden**

Aufgabenstellung

Bedienung mit der Maus:

1. Geben Sie als Zielort „**Apolda**“ ein. Benutzen Sie dazu die Maus und die Funktionen zum Weiter- bzw. Zurückblättern.
2. Geben Sie nun als Zielort „**Neustadt (in Holstein)**“ ein. Benutzen Sie dazu die Filteransicht zur Eingabe der Anfangsbuchstaben.
3. Geben Sie nun über die Filter-Tastatur den Zielort „**Claustal**“ als gesamtes Wort ein. Lassen Sie sich die Liste anzeigen.
4. Korrigieren Sie nun Ihre Eingabe zu „**Clausthal-Zellerfeld**“ und wählen Sie es als Zielort aus.
5. Geben Sie nun als Zielort **einen beliebigen Ort** Ihrer Wahl ein.

Bedienung mittels Sprache:

6. Geben Sie nun als Zielort „**Münster**“ ein. Benutzen Sie dazu ausschließlich Sprachanweisungen und versuchen Sie, ohne Buchstabieren auszukommen.
7. Geben Sie nun (mittels Sprache) „**Ahlen**“ ein.
8. Geben Sie nun den Filter „**Ko**“ ein. Dann lassen Sie sich die Liste aller Städte, die mit „Ko“ beginnen, **vorlesen**. Wählen Sie einen davon aus.
9. Geben Sie nun den Filter „**Ham**“ ein. Lassen Sie sich die Liste anzeigen. Korrigieren Sie sich und geben als Zielort „**Hannover**“ ein.
10. Geben Sie nun wieder **einen beliebigen Ort** Ihrer Wahl ein.

Bedienung nach Belieben:

11. Geben Sie nun den Zielort „**Göppingen**“ ein.
12. Geben Sie nun **so schnell wie möglich** den Zielort „**Bad Homburg**“ ein.
13. Versuchen Sie nun, **ohne dass Sie sich korrigieren müssen**, „**Mülheim**“ als Zielort auszuwählen.
14. Geben Sie nun wieder **einen beliebigen Ort** Ihrer Wahl ein.

Fertig!

Allgemein

Gab es etwas, womit Sie generell nicht zurecht kamen?

☐ ja ☐ nein Wenn ja, was? _____

Modalitätswechsel

Welche Modalität (Sprache oder GUI) haben Sie bevorzugt benutzt?

☐ Sprache ☐ GUI (Graphical User Interface)

Hatten Sie Schwierigkeiten, von einer Modalität zur anderen zu wechseln?

☐ ja ☐ nein

Wenn ja, welcher Art waren die Probleme? (*Mehrfachnennungen möglich*)

- ☐ Das System war zu langsam.
☐ Das System hat meine Eingabe nicht erkannt oder verarbeitet.
☐ Das System hat mir keine Rückmeldung über die Modalität gegeben.
☐ Das System hat von selbst die Modalität gewechselt.
☐ Manchmal war nicht klar, ob nun gerade grafische Eingabe oder Spracheingabe gefordert ist.
☐ Sonstige: _____

PTT-Konzept

War Ihnen das Konzept Push-To-Talk so (oder ähnlich) bereits bekannt?

☐ ja ☐ nein

Wie kamen Sie mit dem PTT-Button zum Aktivieren der Spracherkennung zurecht?

sehr gut	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	sehr schlecht
----------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	---------------

Im Display erschien ein Mikrofon-Symbol, wenn die Spracherkennung aktiviert war.
 War dieses Symbol für Sie als Feedback hilfreich?

sehr hilfreich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gar nicht hilfreich
----------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	---------------------

Ebenso erfolgte ein tonales Signal („Beep“) als Zeichen, dass die Spracherkennung aktiviert war.
 War dieser Ton für Sie als Feedback hilfreich?

sehr hilfreich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gar nicht hilfreich
----------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	---------------------

War Ihnen ab und zu unklar, wann Sie sprechen dürfen und wann nicht?

immer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nie
-------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----

GUI (Graphical User Interface)

Wie haben Sie die Bedienung über die graphische Oberfläche empfunden?

einfach	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	kompliziert
schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	langsam
übersichtlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unübersichtlich
klar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unklar
problemlos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fehleranfällig
direkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	umschweifig
intuitiv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	schwer erlernbar

Sprachdialog

Wie haben Sie die Bedienung mittels Sprache empfunden?

einfach	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	kompliziert
schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	langsam
angenehm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nervig
klar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unklar
problemlos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fehleranfällig
direkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	umschweifig
intuitiv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	schwer erlernbar

Haben Sie eher Kommandos oder kurze Sätze benutzt?

☐ Kommandos ☐ kurze Sätze

Warum? _____

Buchstabieren

Wie haben Sie das Buchstabieren mittels Eigennamen empfunden?

einfach	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	kompliziert
schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	langsam
angenehm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nervig
klar	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unklar
problemlos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fehleranfällig
direkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	umschweifig
intuitiv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	schwer erlernbar

Benutzen Sie das Buchstabieren mittels Eigennamen im Alltag?

☐ täglich ☐ oft ☐ ab und zu ☐ selten ☐ nie

Anhang D

Benutzertest

D.1 Die Versuchsanweisungen

Sehr geehrte(r) Versuchsteilnehmer(in),

zunächst wird Ihnen die Versuchsleiterin die Tracking-Aufgabe mit dem Lenkrad erklären. Danach dürfen Sie diese Aufgabe üben, bis Sie sich darin sicher fühlen und keine Fehler mehr auftreten.

Anschließend wird die Versuchsleiterin Sie über den weiteren Verlauf aufklären.

Viel Spaß und gute „Fahrt“!

Sehr geehrte(r) Versuchsteilnehmer(in),

die Fahraufgabe beherrschen Sie nun sicher. Der nun folgende Teil des Versuchs bezieht sich auf Ihren Dialog mit dem Navigationssystem. Zunächst werden Sie sich nur auf die Zieleingabe konzentrieren, im weiteren Verlauf des Versuchs soll dann parallel die Tracking-Aufgabe bewältigt werden.

Die Versuchsleiterin wird Ihnen die Aufgaben nacheinander vorlesen. Sie können jederzeit nachfragen, wenn Ihnen etwas unklar ist oder sie möchten, dass die Aufgabenstellung wiederholt wird.

Sie haben für jede Aufgabe 4 Minuten Zeit. Während dieser 4 Minuten können Sie den Dialog neu starten und abbrechen, so häufig sie wollen. Ihr Ziel sollte es jedoch sein, die Aufgabe erfolgreich zu beenden. Falls die 4 Minuten abgelaufen sind, wird die Versuchsleiterin sie darauf hinweisen und die nächste Aufgabe vorlesen. Sind Sie bereits früher fertig, geht die Versuchsleiterin sofort zur nächsten Aufgabe über.

Viel Spaß mit den Aufgaben und vorweg schon mal vielen Dank!

Wurde von der Versuchsleiterin VOR den ersten fünf Aufgaben vorgelesen:

Sie bekommen nun zunächst 15 Aufgaben von mir vorgelesen. Davon sollen Sie die ersten fünf ausschließlich mit der Maus bedienen, die nächsten fünf ausschließlich mit Sprache und bei den letzten fünf haben Sie die Wahl.

Ich gebe Ihnen jetzt noch einen kurzen Überblick über das System. Es dient der Eingabe eines Zielortes für ein Navigationssystem. Die Zielführung selber, d.h. Anweisungen des Systems wie „Fahren Sie hier rechts“ oder „In 300 m links abbiegen“ funktioniert leider noch nicht.

Bedienung mit der Maus:

Ihr Ziel ist bei jeder der 15 Aufgaben einen Zielort einzugeben. Dazu müssen Sie sich zunächst die Listenansicht anzeigen lassen, in der Sie den Zielort eingeben können. Die Liste, die Sie dort sehen werden, enthält ca. 1300 Städte innerhalb von Deutschland. In dieser Listenansicht können Sie sich dann vor- bzw. rückwärts bewegen. Sie können einzelne Einträge, d.h. Städte, vorgehen, ganze Seiten oder auch bis ganz an das Ende der Liste. Sie können jederzeit einen Eintrag auswählen.

Bei 1300 Einträgen dauert das Blättern durch die Liste manchmal recht lang. Daher können Sie bei Bedarf in die so genannte Filteransicht wechseln. Dort ist eine Tastatur eingeblendet, mit der Sie einzelne Buchstaben eingeben können. Auf diese Art und Weise können Sie die Anfangsbuchstaben des gewünschten Ortes eingeben und so die Liste einschränken.

Generell gilt: Wenn Sie einmal nicht weiter wissen, probieren Sie einfach etwas aus. Alle Tasten auf dem Display haben auch eine Funktion, versuchen Sie es einfach. Kaputt machen können Sie nichts.

Die Taste „PTT“ ist zum Starten der Spracherkennung. Dazu später mehr. Während der ersten fünf Aufgaben, die Sie nur mit der Maus bedienen sollen, ist diese Taste für Sie tabu.

Wurde von der Versuchsleiterin NACH den ersten fünf Aufgaben vorgelesen:

Bedienung mit Sprache:

Die Taste „PTT“ startet die Spracherkennung. Wenn Sie die Taste gedrückt haben, kommt zunächst immer eine Antwort vom System. Danach können Sie dann sprechen.

Wenn Sie nach ca. 5 Sekunden nichts gesagt haben, schaltet sich die Spracherkennung wieder aus. In diesem Fall drücken Sie einfach noch mal auf „PTT“, dann können Sie genau dort weiter machen, wo Sie aufgehört haben.

Generell versteht das System keine natürlichsprachliche Eingabe mit geschachtelten Sätzen, d.h. komplizierte Formulierungen, wie ich sie jetzt gerade benutze, werden nicht verstanden. Allerdings versteht das System alles von einzelnen Worten zu längeren Sätzen. „Ich möchte den Zielort eingeben“ wird genauso verstanden wie „Zielort“.

Wenn Sie einmal nicht wissen, was Sie sagen sollen, probieren Sie einfach etwas aus. Falls das System Sie trotzdem nicht versteht, liegt das am System und nicht an Ihnen. In diesem Fall formulieren Sie es einfach noch mal anders oder auch kürzer oder länger.

Das Buchstabieren funktioniert leider nicht „normal“, indem man einfach „A B C“ usw. sagt. Statt dessen müssen Sie das rechts eingeblendete Buchstabieralphabet verwenden. D.h. wenn Sie z.B. nach Hildesheim wollen, müssen Sie „Heinrich Ida Ludwig“ sagen.

D.2 Die Aufgabenstellung

Gruppen A1 und B1: Aufgabenstellung 1

Bedienung **ausschließlich** mit der Maus:

1. Geben Sie als Zielort „**Apolda**“ ein. Benutzen Sie dazu die Funktionen zum Weiter- bzw. Zurückblättern.
2. Geben Sie nun die Anfangsbuchstaben „**N e u s t**“ ein.
Lassen Sie sich die **Liste anzeigen**.
Wählen Sie nun als Zielort „**Neustadt (in Holstein)**“ aus.
3. Gehen Sie nun an das **Ende der Liste**.
Wählen Sie dort „**Zwiesel**“ als Zielort aus.
4. Wählen Sie nun die Stadt „**Rotenburg**“ aus, die **an der Fulda** liegt.
5. Geben Sie nun als Zielort **einen beliebigen Ort** Ihrer Wahl ein.

Bedienung **ausschließlich** mittels Sprache:

6. Geben Sie als Zielort „**Münster**“ ein.
7. Geben Sie nun die Anfangsbuchstaben „**K O**“ mit dem Buchstabialphabet ein.
Dann lassen Sie sich die Liste aller Städte, die mit „K O“ beginnen, **vorlesen**.
Wählen Sie einen davon aus.
8. Geben Sie nun als Zielort „**Ahlen**“ (mit H) ein.
9. Wählen Sie nun **einen beliebigen Ort mit „D“** als Zielort aus.
10. Geben Sie nun **Ihren Geburtsort** als Zielort ein.

Bedienung nach Belieben:

11. Geben Sie nun den Zielort „**Göppingen**“ ein.
12. Geben Sie nun den Zielort „**Bad Doberan**“ ein.
13. Suchen Sie nun die Stadt „**Wasserburg**“ aus, die **am Inn** liegt.
14. Geben Sie nun „**Schwarzenberg**“ als Zielort ein.
15. Geben Sie nun **einen Ort** innerhalb von Deutschland ein, **in dem Sie gerne wohnen würden**.

Gruppen A1 und B1: Aufgabenstellung 2

Bedienung *ausschließlich* mit der Maus:

16. Geben Sie als Zielort „**Ansbach**“ ein. Benutzen Sie dazu die Funktionen zum Weiter- bzw. Zurückblättern.
17. Geben Sie nun die Anfangsbuchstaben „**S T O L**“ ein.
Lassen Sie sich die **Liste anzeigen**.
Wählen Sie nun als Zielort „**Stolberg (im Harz)**“ aus.
18. Gehen Sie nun an das **Ende der Liste**.
Wählen Sie dort „**Zwickau**“ als Zielort aus.
19. Wählen Sie nun die Stadt „**Frankfurt**“ aus, die an der polnischen Grenze liegt.
20. Geben Sie nun als Zielort **einen beliebigen Ort** Ihrer Wahl ein.

Bedienung *ausschließlich* mittels Sprache:

21. Geben Sie als Zielort „**Dresden**“ ein.
22. Geben Sie nun die Anfangsbuchstaben „**O B**“ mit dem Buchstabieralphabet ein.
Dann lassen Sie sich die Liste aller Städte, die mit „O B“ beginnen, **vorlesen**.
Wählen Sie einen davon aus.
23. Geben Sie nun als Zielort „**Lahnstein**“ ein.
24. Wählen Sie nun **einen beliebigen Ort mit „M“** als Zielort aus.
25. Geben Sie nun **Ihren Wohnort** als Zielort ein.

Bedienung nach Belieben:

26. Geben Sie nun den Zielort „**Ahrensburg**“ ein.
27. Geben Sie nun den Zielort „**Sankt Augustin**“ ein.
28. Wählen Sie nun **einen beliebigen Ort mit „S“** als Zielort aus.
29. Geben Sie nun „**Brühl**“ als Zielort ein.
30. Geben Sie nun **einen Ort** innerhalb von Deutschland ein, **in dem Sie nicht wohnen möchten**.

Gruppen A1 und B1: Aufgabenstellung 3

Bedienung nach Belieben:

31. Geben Sie nun den Zielort „**Neunkirchen**“ ein.
32. Geben Sie nun den Zielort „**Groß Pankow**“ ein.
33. Wählen Sie nun die Stadt „**Mülheim**“ aus, die **an der Ruhr** liegt.
34. Geben Sie nun „**Buchen**“ als Zielort ein.
35. Geben Sie nun **einen Ort** innerhalb von Deutschland ein, **in dem man mit Dialekt spricht**.

D.3 Der Fragebogen

Sehr geehrte(r) Versuchsteilnehmer(in),

während des Versuchs haben Sie das System näher kennen gelernt. Uns interessiert nun, wie Sie es aus Ihrer persönlichen Sicht einschätzen. Bitte nehmen Sie sich daher kurz die Zeit, die folgenden Fragen zu beantworten. Sie beziehen sich ausschließlich auf das von Ihnen getestete System zur Zieleingabe – die Tracking-Aufgabe sollen Sie nicht bewerten.

Uns interessiert *Ihre* Meinung, d.h. es gibt keine richtigen und falschen Antworten!

Gab es etwas, womit Sie generell nicht zurecht kamen?

- (A1.1) ☐ ja
☐ nein

(A1.2) **Wenn ja, was?** _____

Welche Modalität haben Sie bevorzugt benutzt?

- (B1.1) ☐ Sprache
☐ GUI (Graphical User Interface)
☐ multimodale Bedienung (beides zusammen)

(B1.2) **Warum?** _____

Stellen Sie sich vor, Sie hätten den Zielort während einer Autofahrt eingeben sollen (anstatt der hier realisierten Tracking-Aufgabe). Welche Modalität, glauben Sie, hätten Sie dann bevorzugt?

- (B2.1) ☐ Sprache
☐ GUI
☐ multimodale Bedienung
☐ weiß nicht

Stellen Sie sich vor, die Bedienung der GUI wäre über einen Touchscreen (anstatt der Bedienung mit der Maus) möglich gewesen. Welche Modalität, glauben Sie, hätten Sie dann bevorzugt?

- (B3.1) ☐ Sprache
☐ GUI
☐ multimodale Bedienung
☐ weiß nicht

Wie haben Sie die Bedienung über die graphische Oberfläche empfunden?

C1.1	einfach	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	kompliziert
C1.2	schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	langsam
C1.3	angenehm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nervig
C1.4	verständlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unklar
C1.5	problemlos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fehleranfällig
C1.6	direkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	umschweifig
C1.7	intuitiv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	schwer erlernbar
C1.8	ausgereift	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	verbesserungs- würdig
C1.9	lief locker nebenbei	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	beanspruchte volle Konzentration

Stellen Sie sich vor, die Bedienung der GUI wäre über einen Touchscreen (anstatt der Bedienung mit der Maus) möglich gewesen. Wie, glauben Sie, hätte sich Ihr Verhalten geändert?
(Mehrfachnennungen möglich)

- (C2.1) ☐ Die Bedienung der GUI wäre mir leichter gefallen.
 (C2.2) ☐ Die Bedienung der GUI hätte mir mehr Spaß gemacht.
 (C2.3) ☐ Die Bedienung der GUI hätte weniger Konzentration erfordert.
 (C2.4) ☐ Ich wäre bei der Bedienung der GUI schneller gewesen.
 (C2.5) ☐ Ich hätte bei der Bedienung der GUI weniger Fehler gemacht.

Was hat Ihnen an der GUI-Bedienung am besten gefallen?

(C3.1) _____

Was hat Sie an der GUI-Bedienung am meisten gestört?

(C4.1) _____

Wie haben Sie die Bedienung mittels Sprache empfunden?

D1.1	einfach	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	kompliziert
D1.2	schnell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	langsam
D1.3	angenehm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nervig
D1.4	verständlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unklar
D1.5	problemlos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	fehleranfällig
D1.6	direkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	umschweifig
D1.7	intuitiv	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	schwer erlernbar
D1.8	ausgereift	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	verbesserungs- würdig
D1.9	lief locker nebenbei	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	beanspruchte volle Konzentration

Haben Sie eher längere Sätze oder kurze Kommandos benutzt?

- (D2.1) ☐ Sätze
☐ Kommandos

(D2.2) Warum? _____

Was hat Ihnen an der Sprachbedienung am besten gefallen?

(D3.1) _____

Was hat Sie an der Sprachbedienung am meisten gestört?

(D4.1) _____

Im Laufe Ihres Versuchs wurde ab und zu auf dem Display in der rechten oberen Ecke ein Symbol eingeblendet. Zudem waren in manchen Situationen akustische Signale zu hören.

Beschreiben Sie bitte - in Ihren eigenen Worten - das auf dem Display dargestellte Symbol.

(E1.1) _____

In welchen Situationen und wie lange trat das Symbol auf?

(E1.2) _____

Haben Sie während des Versuchs auf dieses Symbol geachtet?

- (E1.3) ☐ ja
☐ nein

(E1.4) **Wenn ja, welchen Zweck haben Sie damit verfolgt?**

Beschreiben Sie bitte - in Ihren eigenen Worten – das über Lautsprecher ausgegebene akustische Signal.

(E2.1) _____

In welchen Situationen wurde das akustische Signal wiedergegeben?

(E2.2) _____

Haben Sie während des Versuchs auf dieses akustische Signal geachtet?

- (E2.3) ☐ ja
☐ nein

(E2.4) **Wenn ja, welchen Zweck haben Sie damit verfolgt?**

Hätten Sie sich zusätzliche akustische Signale gewünscht?

- (E2.5) ☐ ja
☐ nein

(E2.6) **Wenn ja, in welchen Situationen und warum?**

Rechts sehen Sie nun das Symbol, von dem soeben die Rede war. Es dient zur Statusanzeige der Spracherkennung („an“ bzw. „aus“).

Wenn Ihnen der Ton nicht mehr im Gedächtnis sind, spielt die Versuchsleiterin ihn Ihnen gerne noch einmal vor. Er dient ebenso zur Statusanzeige der Spracherkennung. Dabei ertönte der Ton, sobald die Spracherkennung aktiviert war.



Wie haben Sie das Mikrofon-Symbol zur Statusanzeige der Spracherkennung empfunden?

E3.1	verständlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unklar
E3.2	sehr hilfreich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gar nicht hilfreich
E3.3	angenehm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nervig
E3.4	unauffällig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	aufdringlich

Wie haben Sie das akustische Signal zur Statusanzeige der Spracherkennung empfunden?

E4.1	verständlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	unklar
E4.2	sehr hilfreich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gar nicht hilfreich
E4.3	angenehm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nervig
E4.4	unauffällig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	aufdringlich

Welches Feedback zur Statusanzeige der Spracherkennung fanden Sie insgesamt am besten?

- (E5.1) ☐ visuelles Feedback (Mikrofon-Symbol)
☐ akustisches Feedback (Ton)

War Ihnen ab und zu unklar, wann Sie sprechen dürfen und wann nicht?

E6.1	immer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	nie
------	-------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----

War Ihnen das Konzept zum Starten der Spracherkennung so (oder ähnlich) bereits bekannt?

- (E7.1) ☐ ja
☐ nein

Im Folgenden werden Ihnen einige Aussagen vorgegeben, denen Sie mehr oder weniger zustimmen bzw. die Sie ablehnen können. Bitte kreuzen Sie die Antwortmöglichkeit an, die Sie spontan für die beste halten und die am ehesten auf Sie zutrifft. Bitte beantworten Sie jede Frage!

Wie würden Sie die ergonomische Gestaltung des Systems beurteilen?

		Stimmt gar nicht	Stimmt kaum	Stimmt teilweise	Stimmt über- wiegend	Stimmt völlig
F1.1	Das System war mir unmittelbar verständlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F1.2	Das System gab mir ausreichend Rückmeldung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F1.3	Die Abläufe des Systems waren für mich transparent.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wie würden Sie das System aus Ihrer persönlichen Sicht bewerten?

		Stimmt gar nicht	Stimmt kaum	Stimmt teilweise	Stimmt über- wiegend	Stimmt völlig
F2.1	Dieses System ist sinnvoll.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.2	Mit diesem System würde ich mich wohl fühlen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.3	Dieses System ist nichts für mich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.4	Ich finde dieses System ist sehr nützlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.5	Ich bin von diesem System begeistert.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.6	Diese Technologie im Auto zu besitzen ist für mich ein absolutes Muss.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.7	Ich denke dieses System ist unsicher.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.8	Das System passt zu mir.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.9	Dieses System gefällt mir überhaupt nicht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.10	Dieses System ist eine wertvolle Unterstützung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
F2.11	Das System sagt mir spontan zu.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vielen Dank für Ihre Mühe!

D.4 Der Vorabfragebogen

Robert Bosch GmbH Stuttgart

Abteilung für Informations- und Systemtechnologie

Guten Tag! Vielen Dank für Ihre Mitarbeit im Rahmen eines Versuchs der Robert Bosch GmbH Stuttgart. Das Ausfüllen des Ihnen vorliegenden Fragebogens dauert ca. 10 Minuten. Bitte beantworten Sie alle Fragen in dem dafür vorgesehenen Kästchen.

Bei vielen Fragen können Sie zwischen unterschiedlichen Antwortalternativen wählen. Markieren Sie bitte in einem solchen Fall die für Sie zutreffende Antwort bzw. Aussage mit einem > x <. Ihre Angaben werden gemäss den Datenschutzrichtlinien streng vertraulich behandelt.

Vielen Dank für Ihre Mitarbeit!

1. Persönliche Angaben

- | | | | |
|------|---|----------|-----------------|
| 1.1 | Wie heißen Sie? | Name | Vorname |
| 1.2 | Wann sind Sie geboren (MM.JJJJ)? | | |
| 1.3 | Ihr Geschlecht? | Männlich | Weiblich |
| 1.4 | Was ist Ihre Muttersprache? | | |
| 1.5 | Ihre derzeitige Tätigkeit? | Student | Berufstätig |
| | | | Andere, welche? |
| 1.6 | Welche Art von Tätigkeit bzw. Fach (z.B. Abteilungsleiter oder Biologie)? | | |
| 1.8 | In welchem Alter haben Sie Ihren Führerschein gemacht? | | |
| 1.9 | Haben sie schon einmal an einem ähnlichen Versuch (z.B. im Fahrsimulator) teilgenommen? | Ja | Nein |
| 1.10 | Bitte geben Sie Telefonnummern an, unter denen wir Sie erreichen können. | Tagsüber | Abends |
| | | | Handy |
| 1.11 | Bitte geben Sie Ihre Email-Adresse an, unter der wir Sie am besten erreichen können. | | |

2. Angaben zu Gewohnheiten und gesundheitlichen Einschränkungen

In dem folgenden Abschnitt möchten wir Sie um Angaben zu Ihren Gewohnheiten sowie gesundheitlichen Einschränkungen bitten. Diese Informationen dienen der Vergleichbarkeit der durch die Untersuchung gewonnenen Informationen. Bitte markieren Sie die für Sie zutreffende Antwort mit einem < x >.

- | | | | | |
|-----|--|-------------|--------------|---------------------------|
| 2.1 | Sind Sie Links- oder Rechtshänder? | Linkshänder | Rechtshänder | Beidhänder |
| 2.2 | Benötigen Sie beim Fahren eine Sehhilfe? | Nein | Ja, Brille | Ja, Kontaktlinsen |
| | | | | Kontaktlinsen oder Brille |
| 2.3 | Leiden Sie unter einer Farbsehschwäche? | Nein | Rot/Grün | Blau/Gelb |
| | | | | Andere |
| 2.4 | Leiden Sie unter Hörschäden? | | Ja | Nein |

3. Angaben rund ums Fahrzeug

Im folgenden Abschnitt möchten wir Ihnen ein paar Fragen zu Ihren bisherigen Erfahrungen mit Kraftfahrzeugen stellen.

3.1.1 Was für einen PKW fahren Sie hauptsächlich?

3.1.2 Welches Baujahr hat dieser PKW?

3.2 Seit wieviel Jahren fahren Sie diesen PKW?

3.3 Wieviele Kilometer fahren Sie durchschnittlich im Jahr?
Bitte markieren Sie mit einem > x <.

< 10.000 km

<10.000 km und > 20.000 km

> 20.000 km

3.4 Bitte markieren Sie mit einem > x < welche technischen Systeme sie aktuell in Ihrem PKW haben und wie oft Sie diese nutzen.

		System vorhanden?		Falls vorhanden, wie häufig nutzen Sie diese Technik?				
		ja	nein	sehr selten	selten	teils/teils	häufig	sehr häufig
3.4.1	Navigationssystem							
3.4.2	Tempomat							
3.4.3	Automatische Abstandsregelung (z.B. ACC)							
3.4.4	Parkpilot (Einparkhilfe)							
3.4.5	Telefon							
3.4.6	Radio							
3.4.7	CD und/oder CD-Wechsler							
3.4.8	Klimaanlage							
3.4.9	Bordcomputer							
3.4.10	Multifunktionslenkrad							
3.4.11	Sprachbedienung							
3.4.12	Elektronisches Stabilitätsprogramm							
3.4.13	Automatikgetriebe							
3.4.14	Antiblockiersystem (ABS)							

3.5 Wie sind Ihre bisherigen Erfahrungen mit den von Ihnen genannten Systemen?

Sehr schlecht

Schlecht

Teils teils

Gut

Sehr gut

3.6 Wie hoch würden Sie Ihre Neigung zu neuen Technologien im Auto einschätzen?

Sehr niedrig

Niedrig

Teils teils

Hoch

Sehr hoch

4. Angaben zu technischen Problemen

Im folgenden Abschnitt interessiert uns Ihre Meinung zu Problemen im Umgang mit technischen Geräten.

Es gibt keine richtigen oder falschen Antworten, allein Ihre persönliche Meinung zählt!

Mit „technischen Problemen“ sind im Folgenden Schwierigkeiten im Umgang mit den verschiedensten Geräten aus Alltag und Beruf gemeint, z.B. bei der Programmierung eines Videorecorders, der Arbeit mit dem Computer, der Bedienung einer Mikrowelle, dem Aufstellen von Selbstmontagemöbeln, der Bedienung von Wasserhähnen in öffentlichen Toiletten, dem Lösen von Fahrkarten am Automaten usw. Bitte Markieren Sie den für Sie zutreffenden Fall mit einem > x <.

	Gar nicht	Meist nicht	Weder noch	Zumeist	Absolut
4.1 Ich kann ziemlich viele der technischen Probleme, mit denen ich konfrontiert bin, allein lösen.					
4.2 Technische Geräte sind oft undurchschaubar und schwer zu beherrschen.					
4.3 Es macht mir richtig Spaß, ein technisches Problem zu knacken.					
4.4 Weil ich mit bisherigen technischen Problemen gut zurecht gekommen bin, blicke ich auch künftigen optimistisch entgegen.					
4.5 Ich fühle mich technischen Geräten gegenüber so hilflos, dass ich lieber die Finger von ihnen lasse.					
4.6 Auch wenn Widerstände auftreten, bearbeite ich ein technisches Problem weiter.					
4.7 Wenn ich ein technisches Problem löse, so geschieht das meistens durch Glück.					
4.8 Die meisten technischen Probleme sind so kompliziert, dass es wenig Sinn hat, sich mit ihnen auseinander zusetzen.					

5. Schluss**Vielen Dank für Ihre Mitarbeit!**

Bei eventuellen Fragen oder Problemen wenden Sie sich bitte per Email an isabelle.peters2@de.bosch.com.

Ich wurde darüber informiert, dass meine Angaben gemäß der Datenschutzrichtlinien streng vertraulich behandelt werden und stimme der Speicherung der oben gemachten Angaben zu:

Ja

Nein

Tabellenverzeichnis

3.1	Beispieldialog A	31
3.2	Beispieldialog B	35
3.3	Beispieldialog C	37
3.4	Beispieldialog D	40
4.1	Übersicht GUI-Änderungen	49
4.2	Prompts im Hauptmenü	51
4.3	Prompts in der Listenansicht	51
4.4	Prompts in der Filteransicht	52
4.5	Allgemeine Prompts	53
4.6	Übersicht Sprachdialog-Änderungen	57
5.1	Signifikanz bei Durchschnittlicher Bedienzeit	68
A.1	Deutsches und NATO-Buchstabialphabet	81
A.2	In der Grammatik der Variante definierte Wörter zum Buchstabieren	82

Abbildungsverzeichnis

2.1	Fahrerinformationssystem der AUDI AG	5
2.2	Advanced Traveler Information Systems Configuration	6
2.3	Drei Funktionsbereiche des Advanced Traveler Information System	8
2.4	Architektur eines natürlichsprachlichen Dialogsystems	9
3.1	Das Testsystem in der Übersicht	19
3.2	The model-view-controller design pattern	20
3.3	Die graphische Bedienung über GUI-View/Controller und Application Model	21
3.4	Die sprachliche Bedienung über ASR, NLU, Dialogskript und TTS	22
3.5	Testsystem: Hauptmenü	25
3.6	Testsystem: Fehlermeldung bei Klick auf „Navigation starten“	26
3.7	Testsystem: Listenansicht ohne aktivierten Filter	26
3.8	Testsystem: vergrößertes Kontrollfeld	27
3.9	Testsystem: Filteransicht	28
3.10	Testsystem: Filteransicht mit eingegebenem Filter	28
3.11	Testsystem: Listenansicht mit gefilterter Liste	29
3.12	Testsystem: Listenansicht bei leerer Liste	30
3.13	Testsystem: Hauptmenü mit ausgewähltem Zielort	30
3.14	Testsystem: Hauptmenü mit aktivierter Spracherkennung	32
4.1	Vergleich Testsystem und Variante: Listenansicht	46
4.2	Vergleich Testsystem und Variante: Filteransicht	48
4.3	Vergleich Testsystem und Variante: Hauptmenü	49
4.4	Vergleich Testsystem und Variante: „Aalen“ noch vorhanden bzw. gelöscht	57
5.1	Versuchsaufbau Benutzertest	63
5.2	Tracking-Tool	64

5.3	Task-Completion-Rate	67
5.4	Dialogschritte	67
5.5	Durchschnittliche Bedienzeit über alle Aufgaben	68
5.6	Prozentualer Anteil der drei Blöcke an der Gesamtbedienzeit	69
5.7	Anzahl der „gelben“ Fehler	70
5.8	Dauer der „gelben“ Fehler	70
5.9	Akzeptanz	71
5.10	Negative Aspekte bei der Sprachbedienung	72
5.11	Häufigkeit des Ereignisses „speech-too-early“	73
5.12	Häufigkeit eines Initial Timeout	74
5.13	Häufigkeit einer Äußerung ohne Spracherkenner-Aktivierung	74
5.14	Modalitätswahl zu Beginn der Aufgaben	75
5.15	Anzahl der Modalitätswechsel	76
B.1	Testsystem: Flowchart Legende	83
B.2	Testsystem: Flowchart Hauptmenü	84
B.3	Testsystem: Flowchart Listenansicht	85
B.4	Testsystem: Flowchart Filteransicht	86

Literaturverzeichnis

Bücher, Monographien und Nachschlagewerke

- [Ber98] Niels Ole Bernsen, Hans Dybkjær & Laila Dybkjær (1998): *Designing Interactive Speech Systems: From First Ideas To User Testing*. Berlin: Springer-Verlag.
- [BD02] Jürgen Bortz & Nicola Döring (2002): *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 3. überarb. Auflage. Berlin: Springer-Verlag.
- [Bro04] Felix Brosius (2004): *SPSS 12*. Bonn: mitp-Verlag.
- [BZ05] Achim Bühl & Peter Zöfel (2005): *SPSS 12 - Einführung in die moderne Datenanalyse unter Windows*. 9. überarb. u. erw. Auflage. München: Pearson Studium.
- [Chu97] Gavin E. Churcher & Eric S. Atwell & Clive Souter (1997): *Dialogue Management Systems: a Survey and Overview*. Technical Report 97.06. University of Leeds, School of Computer Studies.
- [Dud00] *Duden - Rechtschreibung* (2000): Dudenredaktion (Hrsg.) 22. Auflage. Mannheim: Dudenverlag.
- [Dud01] *Duden - Das Fremdwörterbuch* (2001): Dudenredaktion (Hrsg.) 7. Auflage. Mannheim: Dudenverlag.
- [Dud04] *Duden - Sinn- und sachverwandte Wörter* (1997): Dudenredaktion (Hrsg.) 2. überarb. Auflage. Mannheim: Bibliographisches Institut & F.A. Brockhaus AG.

- [FHW98] Federal Highway Administration (1998): *Advanced Traveler Information System Capabilities: Human Factors Research Needs*. Publication No. FHWA-RD-98-186. U.S. Department of Transportation, Washington DC, USA. (<http://www.fhwa.dot.gov>)
- [Fla00] David Flanagan (2000): *Java In A Nutshell. Deutsche Ausgabe für Java 1.2 und 1.3*. 3. Auflage. Köln: O'Reilly Verlag.
- [Gib00] Dafydd Gibbon & Inge Mertins & Roger Moore (Eds.) (2000): *Handbook of Multimodal and Spoken Dialogue Systems. Resources, Terminology and Product Evaluation*. CD-ROM. Kluwer Academic Publishers.
- [Gol01] Joachim Goll, Cornelia Weiß & Frank Müller (2001): *Java als erste Programmiersprache. Vom Einsteiger zum Profi*. 3., durchges. u. erw. Auflage. Stuttgart: B.G. Teubner.
- [Kan02] Michaela Kandahle (2002): *Vertrauen Sie Ihrem Butler? Der Zusammenhang von Vertrauen und Akzeptanz bei einem sprachbasierten Assistenzsystem im Kraftfahrzeug*. Diplomarbeit. Humboldt-Universität zu Berlin, Fachbereich Psychologie.
- [Mey02] Jana Meyer (2002): *Entwicklung eines Fragebogens zur Messung der Akzeptanz von innovativen Systemen im Fahrzeug*. Diplomarbeit. Humboldt-Universität zu Berlin, Fachbereich Psychologie.
- [Nie93] Jakob Nielsen (1993): *Usability Engineering*. San Diego: Academic Press.
- [Pra04] Klaas Prause (2004): *Entwicklung von Modulen für eine Spracherkennung*. Diplomarbeit. Fachhochschule Dortmund, Fachbereich Informatik.
- [Rie01] Tobias Rieper (2001): *Spracherkennung für den Zieleingabe-Dialog mit eingebetteten Buchstabiersequenzen und erweitertem Vokabular*. Diplomarbeit, Christian-Albrechts-Universität zu Kiel, Fachbereich Elektro- und Informationstechnik.
- [Sal02] Angelika Salmen (2002): *Multimodale Menüausgabe im Fahrzeug*. Universität Regensburg, Philosophische Fakultät IV (Sprach - und Literaturwissenschaften), Dissertation. München: Herbert Utz Verlag.
- [Sch95] Ernst Günter Schukat-Talamazzini (1995): *Automatische Spracherkennung. Statistische Verfahren der Musteranalyse*. Vieweg Verlag.

- [Shn02] Ben Shneiderman (2002): *User Interface Design*. 3. Auflage. Bonn: mitp-Verlag.
- [Tim00] Klaus-Peter Timpe, Thomas Jürgensohn & Harald Kolrep (Hrsg.) (2000): *Mensch-Maschine-Systemtechnik. Konzepte, Modellierung, Gestaltung, Evaluation*. Düsseldorf: Symposium Publishing GmbH.

Artikel in Zeitschriften oder Sammelbänden

- [Aky01] Suat Akyol & Lars Libuda & Karl-Friedrichs Kraiss (2001): *Multimodale Benutzung adaptiver Kfz-Bordsysteme*. In: Thomas Jürgensohn & Klaus-Peter Timpe (Hrsg.)(2001): *Kraftfahrzeugführung*. Berlin: Springer-Verlag. S. 137-154.
- [Bei99] Guido Beier (1999): *Kontrollüberzeugungen im Umgang mit Technik*. In: *Report Psychologie*, 9/1999, S. 684-693.
- [Chu99] Jennifer Chu-Carroll (1999): *Form-based Reasoning for Mixed-Initiative Dialogue Management in Information-Query Systems*. In: *Proceedings EURO-SPEECH '99*, Vol. 4, S. 1519-1522.
- [Dyb04] Laila Dybkjær, Niels Ole Bernsen & Wolfgang Minker (2004): *Evaluation And Usability Of Multimodal Spoken Language Dialogue Systems*. In: *Speech Communication*, Volume 43, Issues 1-2, June 2004, Elsevier. S. 33-54.
- [Hed00] Volkmar Hedicke (2000): *Multimodalität in Mensch-Maschine-Schnittstellen*. In: [Tim00], S. 203-232.
- [Kel01] Andreas Kellner (2001): *Dialogsysteme*. In: K.-U. Carstensen et al. (Hrsg.): *Computerlinguistik und Sprachtechnologie: eine Einführung*. Heidelberg: Spektrum Akademischer Verlag. S. 484-491.
- [McT02] Michael F. McTear (2002): *Spoken Dialogue Technology: Enabling the Conversational User Interface*. In: *ACM Computing Surveys*, Vol. 34, Issue 1, March 2002, ACM Press. S. 90-169.
- [Top90] William G. Toperzer (1990): *An Advanced Driver Information System Concept and its Communication Considerations*. In: *IEE Communications, Speech and Vision*, Aug. 1994. Vol. 137, Issue 4. S. 364-369.

Internetquellen

- [Aud04] AUDI AG, Ingolstadt. *Audi Deutschland*. Internet, Verifizierungsdatum: 23.12.2004 (<http://www.audi.de>)
- [Bar99] Cristobal Baray (1999): *The model-view-controller (MVC) design pattern*. Internet, Verifizierungsdatum: 11.12.2004 <http://www.cs.indiana.edu/~cbaray/projects/mvc.html>
- [BMV04] BMVBW, Berlin. *Bundesministerium für Verkehr-, Bau- und Wohnungswesen der Bundesrepublik Deutschland*. Internet, Verifizierungsdatum: 13.11.2004 (<http://www.bmwbw.de>)
- [Jav04] Sun Microsystems, Santa Clara, USA. *Java Technology*. Internet, Verifizierungsdatum: 05.10.2004. (<http://www.java.sun.com>)
- [Rea04] Scansoft Inc., USA. *Realspeak Text-to-Speech*. Internet, Verifizierungsdatum: 17.11.2004. (<http://www.realspeak.de>)
- [Sma04] SmartKom, DFKI GmbH, Kaiserslautern. *Dialogische Mensch-Technik-Koordination durch koordinierte Analyse und Generierung multipler Modalitäten*. Internet, Verifizierungsdatum: 12.12.2004. (<http://www.smartkom.org>)
- [Spe04] Speech Experts GmbH, Regensburg. *Gesellschaft für innovative Sprachtechnologie*. Internet, Verifizierungsdatum: 23.10.2004 (<http://www.speech-experts.com>)
- [Sun98] Todd Sundsted (1998): *MVC meets Swing*. JavaWorld, San Francisco, USA. Internet, Verifizierungsdatum: 23.12.2004 (<http://www.javaworld.com/javaworld/jw-04-1998/jw-04-howto.html>)
- [Voi04] Initiative VOICE BUSINESS, Deutschland. *Initiative für ein besseres Image von Sprachdialogsystemen*. Internet, Verifizierungsdatum: 10.12.2004 (<http://www.voiceaward.de>)
- [Wik04] Wikimedia Foundation Inc. *WIKIPEDIA Die freie Enzyklopädie*. Internet, Verifizierungsdatum 17.09.2004 (<http://www.wikipedia.org>)

Ich versichere, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hildesheim, 15.01.2005

Isabelle Peters